DOCUMENT RESUME

ED 057 815 LI 003 304

AUTHOR	Brandhorst, Wesley T.; Marra, Samuel J.; Price, Douglas S.
TITLE	ERIC Data Base; Pagination Field Frequency Distribution Analysis.
INSTITUTION	ERIC Processing and Reference Facility, Bethesda, Md.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
PUB DATE	Nov 71
CONTRACT	OEC-0-70-1494
NOTE	46p.; (0 References)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Data Analysis; *Data Bases; *Information Processing; *Statistical Analysis
IDENTIFIERS	*ERIC Data Base

ABSTRACT

A definitive study of the sizes of documents in the ERIC Data Base is reported by the ERIC Processing and Reference Facility. This is a statistical and frequency distribution study encompassing every item in the file whose record contains pagination data in machine readable form. The study provides pagination data that could be used by present and future EDRS contractors and bidders and, since it is one of the few such comprehensive analyses ever performed on a data base of this size, it provides useful data for the entire information science community. The 36,314 records analyzed constitute virtually the complete ERIC Data Base through June 1971. The data are displayed in two formats: Tables and Graphs. The raw data from which these displays were developed were obtained from detailed computer printouts. (Author/NH)

ED057815

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ERIC PROCESSING AND REFERENCE FACILITY
OPERATED FOR U.S. OFFICE OF EDUCATION *BY* LEASCO SYSTEMS & RESEARCH CORPORATION
4833 RUGBY AVENUE, BETHESDA, MARYLAND 20014

ERIC DATA BASE

PAGINATION FIELD

FREQUENCY DISTRIBUTION ANALYSIS

Contract OEC-0-70-1494, Amendment #6

Final Report on Task 7

November 1971
(Revised)

Prepared by: W. T. Brandhorst, S. J. Marra, D. S. Price

LI 003 304

TABLE OF CONTENTS

I. INTRODUCTION

II. STATEMENT OF WORK

III. METHODOLOGY

A. ED RANGE ANALYZED

B. ACCESSIONS ANALYZED

C. DATA DISPLAY

D. SPECIAL MICROFICHE ANALYSIS

IV. OBSERVATIONS AND CONCLUSIONS

V. TABLES

VI. GRAPHS

LIST OF FIGURES

I. TABLES

A. Regular Tables (Mean/Median/Mode For Each Level)

1. 1966-67
2. 1968
3. 1969
4. 1970
5. 1971 (January-June)
6. 1966-1971 (thru June) - Composite

B. Special Microfiche Analysis (Microfiche Required for EDRS-Available Documents)

7. Level I
8. Level II
9. Level I and II - Composite
10. EDRS-Available Documents Over 1,000 pages in Size

II. GRAPHS

A. Regular Graphs

1. 1966-67 (Level I)
2. 1968 (Level I)
3. " (Level II)
4. " (Level I and II)
5. 1969 (Level I)
6. " (Level II)
7. " (Level III)
8. " (Level I and II)
9. " (Level I, II, and III)
10. 1970 (Level I)
11. " (Level II)
12. " (Level III)
13. " (Level I and II)
14. " (Level I, II, and III)
15. 1971 (Level I)
16. " (Level II)
17. " (Level III)
18. " (Level I and II)
19. " (Level I, II, and III)
20. 1966-71 (thru June) (Level I)
21. " (thru June) (Level II)
22. " (thru June) (Levels I and II)
23. 1969-70 (Level I, II, and III - Pagination Axis Extended)

b. Special Microfiche Analysis

24. Microfiche Required for EDRS-Available Documents
25. Total Accessions Subdivided By Availability Level (I, II, III)

I. INTRODUCTION

One of the more interesting analytical tasks performed by the Facility in 1971 was that concerned with doing a detailed study of the sizes of documents in the ERIC data base.

This topic had arisen earlier in the year in connection with the letting of the contract for the ERIC Document Reproduction Service (EDRS). It was discovered that there was little definitive information which could be given to prospective bidders on which they could base their cost calculations. A few analyses based on samples had been done but these had not allowed for complexities such as: (1) variation over time, i.e., 1967 being different from 1970; (2) variation by type of availability, i.e., Level I documents being different from Level III*; (3) variation by different type of average, i.e., mean being different from median or mode. Just about the only useful piece of data that had come from previous samplings was that everybody agreed the average (mean) document size was somewhere around 70 pages.

It was decided by Central ERIC, therefore, that a definitive study should be made of the sizes of documents in the ERIC data base. This would not be a sample; it would be a statistical and frequency distribution study encompassing every item on the file whose record contained pagination data in machine-readable form. It would settle the size question once and for all; it would provide pagination data that could be used by present and future EDRS contractors and bidders; and it would be one of the few such comprehensive analyses ever performed on a data base of this size, and in that sense would be useful data for the entire information science community.

II. STATEMENT OF WORK

With the above considerations in mind, the following Statement of Work was added to the ERIC Facility contract:

Prepare Frequency Distribution Analysis Report of Pagination Field.

Write, compile, test, de-bug and run computer programs to analyze the pagination field in all ED accessions in the data base to determine frequency of occurrence of reports of all sizes (number of pages). Prepare graphic reports in the general format of Attachment 1 on segments of the file as follows:

(a) By Period - Report for each period as follows: 1966-1967 (combined); 1968; 1969; 1970; 1966-1970.

(b) By Type - For each period listed in (a) above, display pagination frequency for Level I, Level II, Level III, Levels I and II, and total (Levels I, II, III).

*Level I = MF & HC; Level II = MF Only; Level III = Not Available from EDRS

II. STATEMENT OF WORK (CONTINUED)

(c) Calculations - For each of the five (5) curves displayed for each period, calculate and specify:

(1) Mean - Absolute Average, Total pages divided by total documents.

(2) Median - 50th percentile point (half the documents are larger, half are smaller).

(3) Mode - Location of the highest point of the curve.

III. METHODOLOGY

A. ED Range Analyzed

The ED range analyzed was that which applies to Research in Education. It extends from ED-010000, the first number in the first issue of RIE (November 1966), to ED-048516, the last number in the June 1971 RIE.

It was decided not to analyze the early non-RIE collections, e.g., Disadvantaged (ED-001001 - 002740 = 1,740 accessions) and Research Reports 1956-1965 (ED-002747 - 003960 = 1,214 accessions) for the following reasons:

1. The Disadvantaged collection is a specialized collection of documents and does not cover the entire spectrum of the literature as does RIE. It may, therefore, be atypical in certain respects, including possibly pagination, and should not be included for this reason.

2. The pre-RIE records are less reliable in their content and format then RIE records, having been converted from some earlier machine-readable format. They present more problems to the program (variables and special cases) than the RIE records and would result in a higher percentage of exception records (printed out by the program as unprocessable).

3. The focus of the study is RIE. It would provide a more concise and more easily remembered explanation if one could state that the study simply covers RIE 1966-1971.

It should be said, however, that the pagination analysis programs could easily be run against this early non-RIE data (2,954 accessions) should OE desire at any time in the future to do so. We merely felt that for purposes of this study there was adequate justification for excluding them.

III. METHODOLOGY (CONTINUED)

B. Accessions Analyzed

Between ED-010000 and ED-048516 there are a theoretical total of 38,516 records that could have been considered for statistical analysis. The actual number of records that were considered by the programs was 36,314. The 2,202 RIE records that were not considered consist of the following kinds of records:

1. AIM and ARM Collections - These items consist of collections of documents under one ED number. These collections sometime amount to hundreds of documents with total paginations up to 24,000 pages. If included they would have artificially and wrongfully skewed the distribution and the averages to a significant extent.

2. Non-Print Items - records, tapes, films, etc. There are very few such items in the ERIC data base. Where they appear they generally lack pagination data.

3. Accessions Bearing No Pagination Data

These are relatively rare in all years except 1968 when there appears to have been a policy of not paginating Level III documents. No Level III documents in 1968 have pagination data in their records. This is the largest category of documents not analyzed, comprising over 90% of the total amount not considered.

4. Accessions Bearing Garbled Pagination Data

The pagination data in ERIC records appears in the Descriptive Note field. Other types of data may also appear in this field. When this occurs the data fields are separated by a semi-colon. In order to zero in on the pagination figure and to distinguish it from the other data in the field it was necessary to test for a distinguishing characteristic. This characteristic was the "p." which generally appears after the numeric pagination value. In some few cases no "p." existed; either one or both of the characters were missing. A record containing this flaw was printed out as an exception record and was not analyzed. These were relatively few compared to the accessions bearing no pagination data.

III. METHODOLOGY (CONTINUED)

B. Accessions Analyzed (Continued)

5. Accessions Missing From the File

The count of 1968 records on the ERIC Resume Master File turned out to be 15 less than what it should theoretically have been, indicating 15 missing records. All other years checked out exactly. This validation of file content was an unexpected dividend of the pagination analysis work.

In summary, of the 38,516 records that we theoretically started with, a total of 2,202 (or 5%) were eliminated as "defective" in the various ways enumerated above, and were not permitted to impact the statistics presented in the tables and graphs of this report. The vast majority of the "defective" records were 1968 Level III documents, containing no pagination data to analyze.

The 36,314 records analyzed constitute virtually the complete ERIC Data Base through June 1971 and are obviously more than a sufficient base from which to draw generalizations.

C. Data Display

The data which constitute this report are displayed in two formats: Tables and Graphs, in order to accommodate the varying preferences of the many different individuals who will be making use of the figures. The Tables support the Graphs in the sense that they present many of the specific numbers represented by points on the graph lines. The Graphs support the Tables in the sense that they show, in one easily grasped picture, what the numbers are doing, how they are moving, what the trends are, etc. Behind both Tables and Graphs are the detailed computer printouts representing the raw data from which both displays were developed. The computer printouts will be maintained in storage indefinitely in order to obviate the need to ever run again the early part of the file. It is envisaged that future runs will probably be made at the end of each year to cover that year.

1. Tables

The Tables are all arranged in the same basic matrix format. Down the left hand side there are arranged the various "Levels", i.e., I, II, III, I and II (the "availables"), I and II and III (the composite). Across the top are arranged the various types of averages, i.e., mode, mean,

III. METHODOLOGY (CONTINUED)

C. Data Display (Continued)

1. Tables (Continued)

median. Also provided are figures for the total number of documents at the mode value, the total documents at each "Level", and the total pages at each "Level". In other words, virtually everything that anyone could ask about pagination has been presented.

There are tables for each calendar year and each page is labeled with the accession range that occurred in that year. Because of the small number of documents announced in 1966 (two RIE issues), 1966 and 1967 data were grouped together. A super-composite table, covering the years 1966-1971 (June), presents the data for the entire 36,314 records analyzed, and is, in effect, all the other tables combined into one.

2. Graphs

Each Table has its corresponding graph. It was found that each graph falls off very rapidly toward zero after you pass 220 pages, but each graph also strings out in very attenuated fashion up to some very high page values. For example, the 1970 distribution ranges from 1 - 3261 pages but 88.95 percent of the documents are below 220 pages. This presented a problem in display, as the horizontal axis threatened to completely dominate the vertical axis. Since after 220 pages the graph lines are all essentially flat and "hugging" the horizontal axis, it was finally decided to stop each graph at that value and provide the percentile reached at that point. This treatment would allow reasonable pagination ranges to be used on the horizontal axis and would prevent a compression effect to the left that would have made the graphs much less meaningful and informative. To demonstrate the effect of this kind of distribution on a display, we then decided to do one graph that would encompass the entire pagination range of a given period of time without truncation. We decided upon the two most recent full years, 1969 - 1970. It is easy to see, from an examination of this last graph, why the decision to stop the regular graphs at 220 pages was a reasonable compromise.

In addition to its basic curve, each graph comes equipped with a small inset table summarizing the most significant values shown on the graph. This table also presents the pagination range dealt with by the graph and calculates the 75% percentile and the percentile at 220 pages, where the graph truncates. These inset tables will in almost all cases make it unnecessary to refer back to the 'Table Only" section.

III. METHODOLOGY (CONTINUED)

C. Data Display (Continued)

2. Graphs (Continue)

Vertical lines have been drawn on each graph representing the range mode, the mean, the median (50th percentile), and the 75th percentile. A word is perhaps in order concerning the range mode. The mode is that page value that is most frequently represented in the collection. We call this the "individual mode". Since, however, for graphing purposes it was necessary to select pagination ranges to plot points against (rather than having each individual page value represented on the horizontal axis), we ended up with a pagination range that was the most frequently represented in the collection. We called this the "range mode". In the vast majority of cases the individual mode fell within the range mode, but not always. For example, in 1970 the overall individual mode was 10 pages (where there were 212 documents). However, the range mode worked out to be 11-20 pages. This is a minor irritant however, for it takes very little work with these tables and graphs to become convinced that the mode is probably the least useful of the averages calculated. It is of curiosity value only, for instance, that in all the graphs there was only one bi-modal distribution. 1971 Level III documents had two individual modes: 21-30 pages and 121-130 pages, each range having 65 documents representing it.

D. Special Microfiche Analysis

The median for the combined group of EDRS-available documents (Level I and II) is surprisingly low (39). This suggests that an unexpectedly large number of ERIC documents will fit on one fiche and raises the interesting question of the relationship between reduction ratio (i.e., in effect the number of pages that will fit on one fiche) and the percentage of the collection that will fit on one fiche. We have not explored this subject completely but have restricted ourselves to calculating exactly what percentile of the ERIC collection it is that will fit on one fiche, two fiche, three fiche, etc., at the existing reduction ratio. The special Table presenting this data is organized on the vertical axis by "cards/pages", i.e., 1 microfiche bearing documents 1-57 pages in size, 2 microfiche bearing documents 58-127 pages in size, etc., and across the horizontal axis by year and then by number of documents at each card/pages level (with running cumulative percentages). The far right columns present the composite values for 1966-1971.

This table, and its associated graph, undoubtedly has the most to say to the EDRS contractor. From this data, for example, it can be seen that 61.82% of the total EDRS-available documents will fit on

III. METHODOLOGY (CONTINUED)

D. Special Microfiche Analysis (Cont.)

one fiche; 83.50% of the total EDRS-available documents will fit on two fiche or less, and so on. By six fiche the graph is essentially on a plateau as only 1.57% of the documents require more than six fiche.

As byproducts of the special microfiche analysis, a table has been prepared listing all documents in the data base over 1,000 pages in size, and a graph has been prepared subdividing each year's accessions by "Availability Level", i.e., Level I, II, or III.

IV. OBSERVATIONS AND CONCLUSIONS

Each statistician will have his own set of observations about the data revealed in this report. The data is now available for many analyses and extrapolations. We feel that study of this data will cause many useful facts to emerge.

Since the task of this study was simply to obtain the data and to present it in frequency distribution format, we have carefully restrained our strong desire to generalize. Nevertheless, some basic observations are in order:

Mean - The earlier ideas about the nature of the mean are pretty well born out. For documents that are available from EDRS in hard copy (Level I), it is around 70 pages (1968 = 70, 1969 = 75, 1970 = 68, 1971 = 68). For documents that are available only in microfiche (Level II), however, it goes up sharply (1968 = 135, 1969 = 115, 1970 = 110, 1971 = 104). For documents that are not available at all from EDRS (Level III) the rise is even farther (1968 = 215, 1969 = 152, 1970 = 174, 1971 = 180).

Shape of Curve - The distribution is not a normal or bell-shaped distribution. It is greatly elongated to the right hand side. Some elongation was suspected but the extent of it came as a surprise. The effect of it is, of course, to pull the median and the mean away from the mode. The mean is much more affected by the presence of large "giant" documents than is the median, and is pulled far to the right as a result.

Mode - The smallness of the mode value was surprising to us. We anticipated a value closer to 50 pages, but the mode is almost universally in the 11-20 range. As we said before, the general peak of the curve (or range mode) is interesting but the individual mode is one of the least interesting and most "accidental" figures.

IV. OBSERVATIONS AND CONCLUSIONS (CONTINUED)

Median - The median was also smaller than we had originally anticipated. For Level I documents it runs around 35 pages (1968 = 36, 1969 = 38, 1970 = 34, 1971 = 36). For Level II documents it generally runs just over 60 pages (1968 = 76, 1969 = 68, 1970 = 55, 1971 = 60). When these figures are combined (1968 = 38, 1969 = 40, 1970 = 37, 1971 = 39), the median for the combined group of EDRS-available documents is surprising low, around 39 pages. It is the median, and its associated percentile figures that have been converted into the "special microfiche analysis" and that have the most to say to the EDRS contractor.

In closing we would like to suggest that the next logical extension of this investigation is to match its data against order data from EDRS. Are the document orders equally spread (i.e., normally distributed) across the documents available, or do they restrict themselves to certain sizes and therefore price ranges?

FREQUENCY DISTRIBUTION REPORT TOTALS

DATA FOR YEAR(S)--1966-67

ACCESSION NUMBER RANGE-ED010000-ED012348

LEVEL TYPE	MODE PAGE	MODE DOCUMENTS	TOTAL DOCUMENTS	TOTAL PAGES	MEAN PAGE	MEDIAN PAGE
LEVEL I	00014	00044	02206	000199031	00090	00049
LEVEL II	00051	00001	00004	000000353	00088	00062
LEVEL III	00018	00001	00002	000000041	00021	00020
LEVEL I AND II	00014	00044	02210	000199384	00090	00049
LEVELS I II AND III	00014	00044	02212	000199425	00090	00049

TABLE 1

FREQUENCY DISTRIBUTION REPORT TOTALS

DATA FOR YEAR(S)--1968

ACCESSION NUMBER RANGE-ED012349-ED021151

LEVEL TYPE	MODE PAGE	MODE DOCUMENTS	TOTAL DOCUMENTS	TOTAL PAGES	MEAN PAGE	MEDIAN PAGE
LEVEL I	00016	00150	06668	000464650	00070	00036
LEVEL II	00035	00007	00274	000037016	00135	00076
LEVEL III	00215	00001	00001	000000215	00215	00
LEVELS I AND II	00016	00150	06942	000501666	00072	00038
LEVELS I II AND III	00016	00150	06943	000501881	00072	00038

TABLE 2

FREQUENCY DISTRIBUTION REPORT TOTALS

DATA FOR YEAR(S)--1969

ACCESSION NUMBER RANGE-ED021152-ED031604

LEVEL TYPE	MODE PAGE	MODE DOCUMENTS	TOTAL DOCUMENTS	TOTAL PAGES	MEAN PAGE	MEDIAN PAGE
LEVEL I	00014	00166	07592	000570102	00075	00038
LEVEL II	00036	00013	00624	000070374	00113	00068
LEVEL III	00004	00035	02091	000316906	00152	00127
LEVELS I AND II	00014	00170	08216	000640476	00078	00040
LEVELS I II AND III	00007	00188	10307	000957382	00093	00048

TABLE 3

FREQUENCY DISTRIBUTION REPORT TOTALS

DATA FOR YEAR(S)--1970

ACCESSION NUMBER RANGE-ED031605-ED042060

LEVEL TYPE	MODE PAGE	MODE DOCUMENTS	TOTAL DOCUMENTS	TOTAL PAGES	MEAN PAGE	MEDIAN PAGE
LEVEL I	00010	00179	07204	000486884	00068	00034
LEVEL II	00024	00020	01213	000131233	00108	00055
LEVEL III	00064	00029	02004	000349561	00174	00136
LEVELS I AND II	00010	00196	08417	000618117	00073	00037
LEVELS I II AND III	00010	00212	10421	000967678	00093	00046

TABLE 4

FREQUENCY DISTRIBUTION REPORT TOTALS

DATA FOR YEAR(S)-- JAN-JUN 1971

ACCESSION NUMBER RANGE-ED042061-ED048516

LEVEL TYPE	MODE PAGE	MODE DOCUMENTS	TOTAL DOCUMENTS	TOTAL PAGES	MEAN PAGE	MEDIAN PAGE
LEVEL I	00013	00119	04444	000302351	00068	00036
LEVEL II	00008	00012	00627	000065020	00104	00060
LEVEL III	00016	00012	01360	000244245	00180	00143
LEVELS I AND II	00013	00124	05071	000367371	00072	00039
LEVELS I II AND III	00013	00127	06431	000611616	00095	00050

TABLE 5

FREQUENCY DISTRIBUTION REPORT TOTALS

DATA FOR YEAR(S) -- 1966-JUN 1971

ACCESSION NUMBER RANGE-ED010000-ED048516*

LEVEL TYPE	MODE PAGE	MODE DOCUMENTS	TOTAL DOCUMENTS	TOTAL PAGES	MEAN PAGE	MEDIAN PAGE
LEVEL I	00011	00602	28114	002023018	00072	00037
LEVEL II	00024	00044	02742	000303996	00111	00061
LEVELS I AND II	00011	00625	30856	002327014	00075	00039

*EDRS-Available (LEVELS I & II) Only.

TABLE 6

ERIC DATA BASE

PAGINATION FIELD ANALYSIS

MICROFICHE REQUIRED FOR EDRS-AVAILABLE DOCUMENTS (LEVEL 1)

	1966-1967				1968				1969				1970				1971				TOTAL			
CARDS/ PAGES	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT
1 1-57	1,225	55.53	1,225	55.53	4,265	63.96	4,265	63.96	4,713	62.08	4,713	62.08	4,693	65.14	4,693	65.14	2,858	64.31	2,858	64.31	17,754	63.15	17,754	63.15
2 58-127	509	23.07	1,734	78.60	1,379	20.68	5,644	84.64	1,678	22.10	6,391	84.18	1,493	20.73	6,186	85.87	935	21.04	3,793	85.35	5,994	21.32	23,748	84.47
3 128-197	216	9.79	1,950	88.39	525	7.88	6,169	92.52	628	8.27	7,019	92.45	483	6.70	6,669	92.57	358	8.06	4,151	93.41	2,210	7.86	25,958	92.33
4 198-267	95	4.31	2,045	92.70	222	3.33	6,391	95.85	261	3.44	7,280	95.89	258	3.58	6,927	96.15	125	2.81	4,276	96.22	961	3.42	26,919	95.75
5 268-337	62	2.81	2,107	95.51	129	1.93	6,520	97.78	123	1.62	7,403	97.51	124	1.72	7,051	97.87	85	1.91	4,361	98.13	523	1.86	27,442	97.61
6 338-407	45	2.04	2,152	97.55	52	0.78	6,572	98.56	64	0.84	7,467	98.35	70	0.97	7,121	98.84	35	0.79	4,396	98.92	266	0.95	27,708	98.56
7 408-477	24	1.09	2,176	98.64	37	0.55	6,609	99.11	38	0.50	7,505	98.85	23	0.32	7,144	99.16	17	0.38	4,413	99.30	139	0.49	27,847	99.05
8 478-547	15	0.68	2,191	99.32	30	0.45	6,639	99.56	22	0.29	7,527	99.14	22	0.31	7,166	99.47	9	0.20	4,422	99.50	98	0.35	27,945	99.40
9 548-617	8	0.36	2,199	99.68	13	0.20	6,652	99.76	5	0.07	7,532	99.21	12	0.17	7,178	99.64	5	0.11	4,427	99.61	43	0.15	27,988	99.55
10 618-687	4	0.18	2,203	99.86	4	0.06	6,656	99.82	18	0.24	7,550	99.45	9	0.13	7,187	99.77	7	0.16	4,434	99.77	42	0.15	28,030	99.70
11 688-757	0	0.00	2,203	99.86	4	0.06	6,660	99.88	8	0.10	7,558	99.55	6	0.08	7,193	99.85	3	0.07	4,437	99.84	21	0.07	28,051	99.77
12 758-827	0	0.00	2,203	99.86	4	0.06	6,664	99.94	6	0.08	7,564	99.63	3	0.04	7,196	99.89	2	0.05	4,439	99.89	15	0.05	28,066	99.82
13 828-897	0	0.00	2,203	99.86	0	0.00	6,664	99.94	6	0.08	7,570	99.71	1	0.01	7,197	99.90	0	0.00	4,439	99.89	7	0.03	28,073	99.85
14 898-967	2	0.09	2,205	99.95	1	0.01	6,665	99.95	2	0.03	7,572	99.74	1	0.01	7,198	99.91	1	0.02	4,440	99.91	7	0.03	28,080	99.88
15 & Over 968-up	1	0.05	2,206	100.00	3	0.05	6,668	100.00	20	0.26	7,592	100.00	6	0.09	7,204	100.00	4	0.09	4,444	100.00	34	0.12	28,114	100.00

TABLE 7

PAGINATION FIELD ANALYSIS

DOCUMENTS PER MICROFICHE

LEVEL II

	1966-1967				1968				1969				1970				1971				TOTAL			
CARDS/ PAGES	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT
1 1-57	2	50.00	2		100	36.50	100	36.50	276	44.23	276	44.23	639	52.68	639	52.68	304	48.48	304	48.48	1,321	48.18	1,321	48.18
2 58-127	1	25.00	3	75.00	75	27.37	175	63.87	171	27.40	447	71.63	279	23.00	918	75.68	170	27.11	474	75.59	696	25.38	2,017	73.56
3 128-197	1	25.00	4	100.00	39	14.23	214	78.10	80	12.82	527	84.45	126	10.39	1,044	86.07	60	9.57	534	85.16	306	11.16	2,323	84.72
4 198-267	0	0	4	100.00	25	9.12	239	87.22	36	5.77	563	90.22	81	6.68	1,125	92.75	29	4.63	563	89.79	171	6.24	2,494	90.96
5 268-337	0	0	4	100.00	19	6.94	258	94.16	26	4.17	589	94.39	32	2.64	1,157	95.39	29	4.63	592	94.42	106	3.86	2,600	94.82
6 338-407	0	0	4	100.00	8	2.92	266	97.08	18	2.89	607	97.28	22	1.81	1,179	97.20	15	2.39	607	96.81	63	2.30	2,663	97.12
7 408-477	0	0	4	100.00	4	1.46	270	98.54	4	0.64	611	97.92	10	0.82	1,189	98.02	7	1.11	614	97.92	25	0.91	2,688	98.03
8 478-547	0	0	4	100.00	1	0.36	271	98.90	3	0.48	614	98.40	4	0.33	1,193	98.35	2	0.32	616	98.24	10	0.36	2,698	98.39
9 548-617	0	0	4	100.00	0	0.00	271	98.90	5	0.80	619	99.20	5	0.41	1,198	98.76	4	0.64	620	98.88	14	0.51	2,712	98.90
10 618-687	0	0	4	100.00	0	0.00	271	98.90	0	0.00	619	99.20	0	0.00	1,198	98.76	4	0.64	624	99.52	4	0.15	2,716	99.05
11 688-757	0	0	4	100.00	1	0.37	272	99.27	0	0.00	619	99.20	3	0.25	1,201	99.01	1	0.16	625	99.68	5	0.18	2,721	99.23
12 758-827	0	0	4	100.00	1	0.36	273	99.63	2	0.32	621	99.52	2	0.17	1,203	99.18	1	0.16	626	99.84	6	0.22	2,727	99.45
13 828-897	0	0	4	100.00	0	0.00	273	99.63	1	0.16	622	99.68	1	0.08	1,204	99.26	1	0.16	627	100.00	3	0.11	2,730	99.56
14 898-967	0	0	4	100.00	0	0.00	273	99.63	0	0.00	622	99.68	1	0.08	1,205	99.34	0	0	627	100.00	1	0.04	2,731	99.60
15 & Over 968-Up	0	0	4	100.00	1	0.37	274	100.00	2	0.32	624	100.00	8	0.66	1,213	100.00	0	0	627	100.00	11	0.40	2,742	100.00

TABLE 8

PAGINATION FIELD ANALYSIS

DOCUMENTS PER MICROFICHE

LEVELS I AND II

	1966-1967				1968				1969				1970				1971				TOTAL			
CARDS/ PAGES	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT	NO. OF DOCS.	PER-CENT	CUM. DOCS.	CUM. PER-CENT
1 1-57	1,227	55.52	1,227	55.52	4,365	62.88	4,365	62.88	4,989	60.72	4,989	60.72	5,332	63.35	5,332	63.35	3,162	62.36	3,162	62.36	19,075	61.82	19,075	61.82
2 58-127	510	23.08	1,737	78.60	1,454	20.94	5,819	83.82	1,849	22.51	6,838	83.23	1,772	21.05	7,104	84.40	1,105	21.79	4,267	84.15	6,690	21.68	25,765	83.50
3 128-197	217	9.82	1,954	88.42	564	8.13	6,383	91.95	708	8.62	7,546	91.85	609	7.24	7,713	91.64	418	8.24	4,685	92.39	2,516	8.15	28,281	91.65
4 198,267	95	4.30	2,049	92.72	247	3.56	6,630	95.51	297	3.61	7,843	95.46	339	4.02	8,052	95.66	154	3.03	4,839	95.42	1,132	3.67	29,413	95.32
5 268-337	62	2.80	2,111	95.52	148	2.13	6,778	97.64	149	1.81	7,992	97.27	156	1.86	8,208	97.52	114	2.25	4,953	97.67	629	2.04	30,042	97.36
6 338-407	45	2.04	2,156	97.56	60	0.86	6,838	98.50	82	1.00	8,074	98.27	92	1.09	8,300	98.61	50	0.99	5,003	98.66	329	1.07	30,371	98.43
7 408-477	24	1.08	2,180	98.64	41	0.59	6,879	99.09	42	0.51	8,116	98.78	33	0.39	8,333	99.00	24	0.47	5,027	99.13	164	0.53	30,535	98.96
8 478-547	15	0.68	2,195	99.32	31	0.45	6,910	99.54	25	0.31	8,141	99.09	26	0.31	8,359	99.31	11	0.22	5,038	99.35	108	60.35	30,643	99.31
9 548-617	8	0.36	2,203	99.68	13	0.19	6,923	99.73	10	0.12	8,151	99.21	17	0.20	8,376	99.51	9	0.17	5,047	99.52	57	0.18	30,700	99.49
10 618-687	4	0.18	2,207	99.86	4	0.06	6,927	99.79	18	0.22	8,169	99.43	9	0.11	8,385	99.62	11	0.22	5,058	99.74	46	0.15	30,746	99.64
11 688-757	0	0.00	2,207	99.86	5	0.07	6,932	99.86	8	0.10	8,177	99.53	9	0.11	8,394	99.73	4	0.08	5,062	99.82	26	0.09	30,772	99.73
12 758-827	0	0.00	2,207	99.86	5	0.07	6,937	99.93	8	0.10	8,185	99.63	5	0.06	8,399	99.79	3	0.06	5,065	99.88	21	0.07	30,793	99.80
13 828-897	0	0.00	2,207	99.86	0	0.00	6,937	99.93	7	0.08	8,192	99.71	2	0.02	8,401	99.81	1	0.02	5,066	99.90	10	0.03	30,803	99.83
14 898-967	2	0.09	2,209	99.95	1	0.01	6,938	99.94	2	0.02	8,194	99.73	2	0.02	8,403	99.83	1	0.02	5,067	99.92	8	0.03	30,811	99.86
15 & Over 968-Up	1	0.05	2,210	100.00	4	0.06	6,942	100.00	22	0.27	8,216	100.00	14	0.17	8,417	100.00	4	0.08	5,071	100.00	45	0.14	30,856	100.00

TABLE 10

EDRS - AVAILABLE DOCUMENTS OVER 1,000 PAGES*

LEVEL I

NO.	PAGES	NO.	PAGES	NO.	PAGES
ED010106	3,141	ED024023	1,717	ED026546	1,232
ED010408	1,423	ED024925	1,554	ED026635	2,943
ED013197	1,137	ED024936	1,011	ED027369	1,201
ED015677	1,596	ED024937	1,326	ED032546	1,567
ED018676	1,287	ED024942	1,417	ED036281	1,352
ED021206	2,140	ED024943	1,451	ED040303	1,041
ED022178	1,114	ED024955	1,265	ED041344	3,261
ED022179	2,707	ED024956	1,545	ED041685	1,143
ED023083	2,053	ED024957	1,428	ED042728	1,044
ED023094	1,750	ED025742	2,645	ED044757	1,371
ED023096	1,071	ED025774	1,477	ED047798	1,262

LEVEL II

NO.	PAGES	NO.	PAGES	NO.	PAGES
ED019980	3,069	ED031697	1,643	ED034203	1,720
ED030886	2,412	ED032544	1,268	ED034471	1,912
ED031461	1,451	ED033357	3,178	ED038315	1,322
ED031688	3,245	ED034196	3,008		

*Through June 1971. Excludes AIM/ARM collections.

ERIC DATA BASE PAGINATION FIELD ANALYSIS

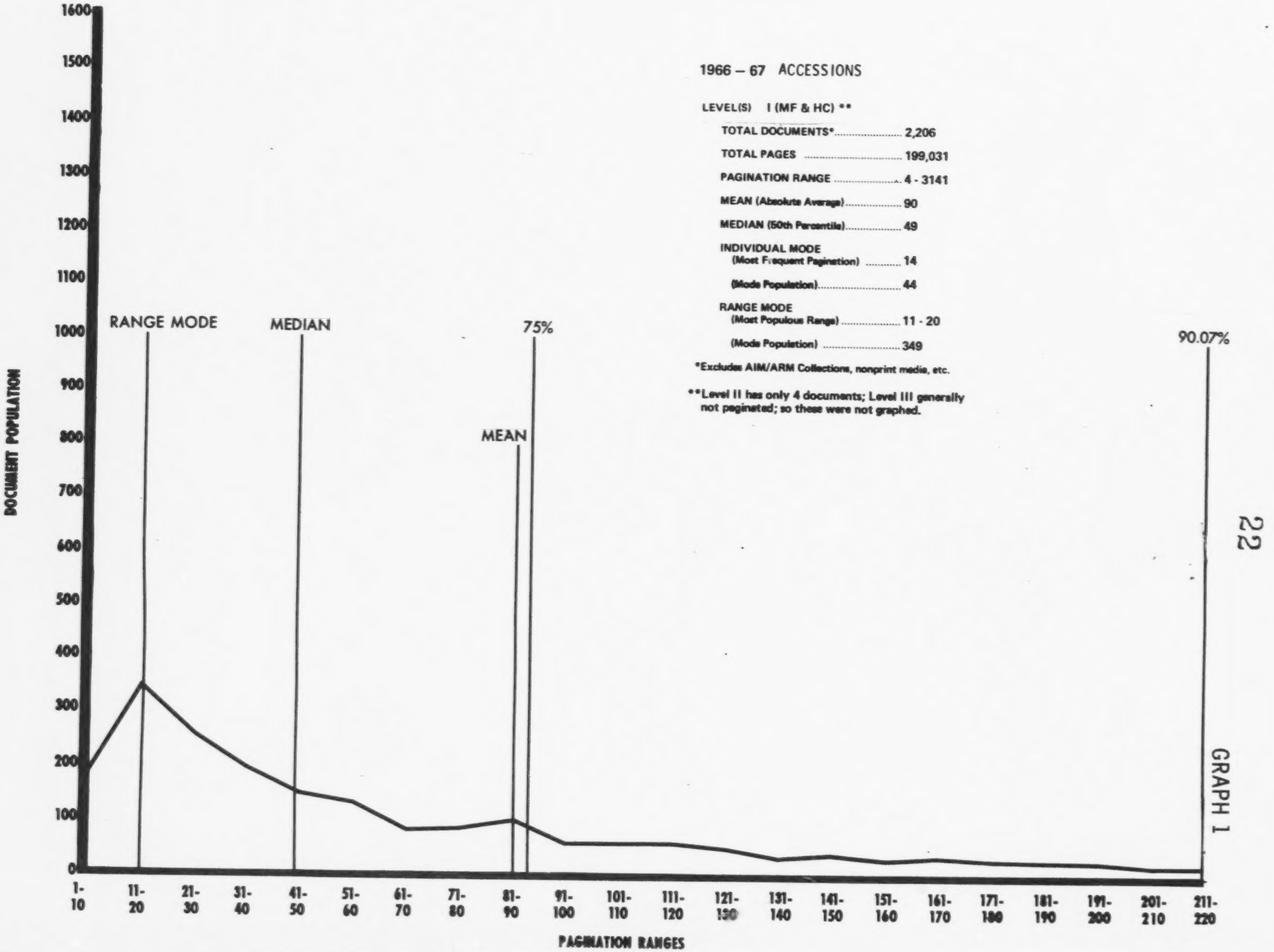
1966 – 67 ACCESSIONS

LEVEL(S) I (MF & HC) **

TOTAL DOCUMENTS*	2,206
TOTAL PAGES	199,031
PAGINATION RANGE	4 - 3141
MEAN (Absolute Average)	90
MEDIAN (50th Percentile)	49
INDIVIDUAL MODE (Most Frequent Pagination)	14
(Mode Population)	44
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	349

*Excludes AIM/ARM Collections, nonprint media, etc.

**Level II has only 4 documents; Level III generally not paginated; so these were not graphed.



GRAPH 1

ERIC DATA BASE PAGINATION FIELD ANALYSIS

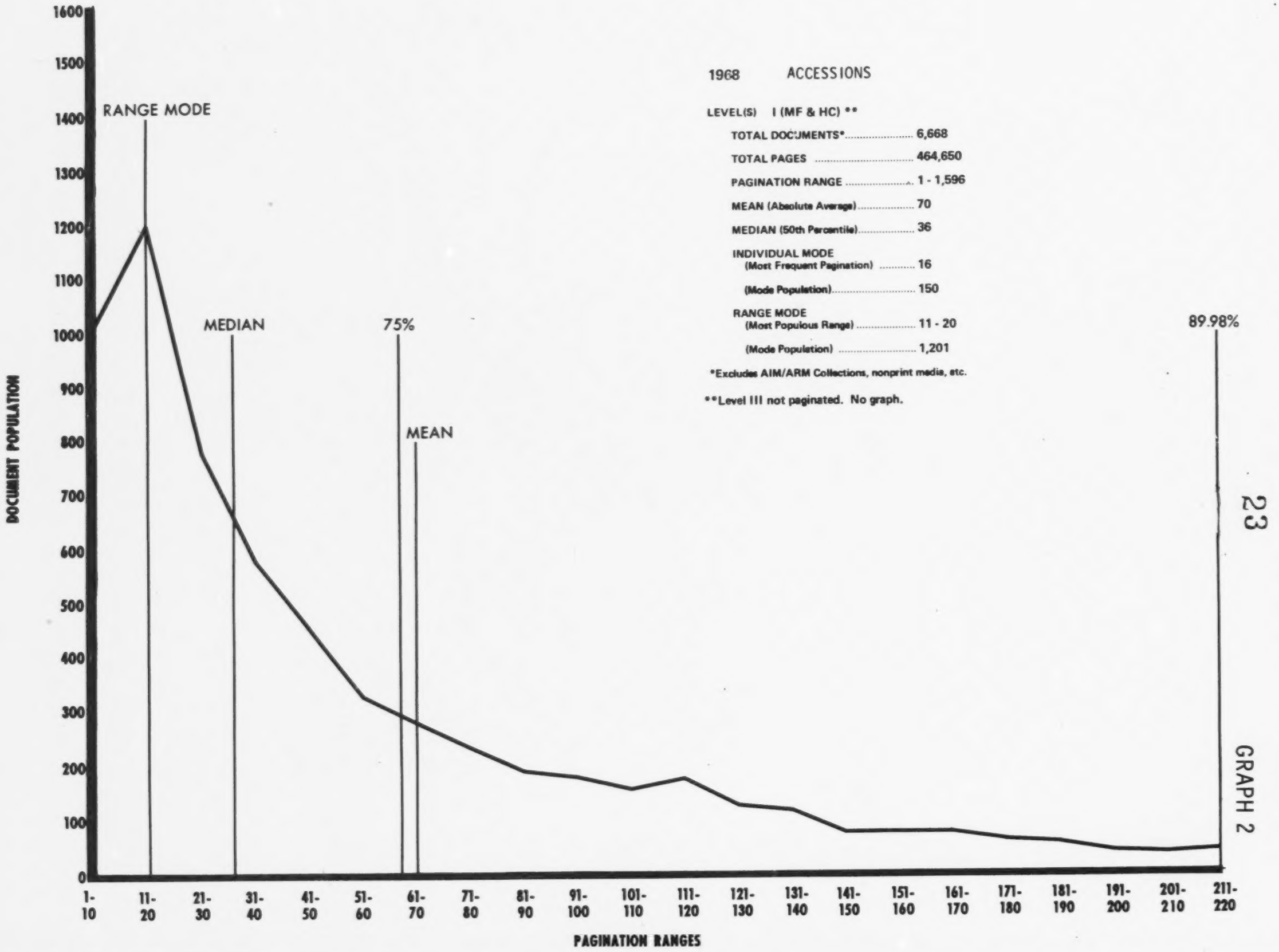
1968 ACCESSIONS

LEVEL(S) I (MF & HC) **

TOTAL DOCUMENTS*	6,668
TOTAL PAGES	464,650
PAGINATION RANGE	1 - 1,596
MEAN (Absolute Average)	70
MEDIAN (50th Percentile)	36
INDIVIDUAL MODE (Most Frequent Pagination)	16
(Mode Population)	150
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	1,201

*Excludes AIM/ARM Collections, nonprint media, etc.

**Level III not paginated. No graph.



GRAPH 2

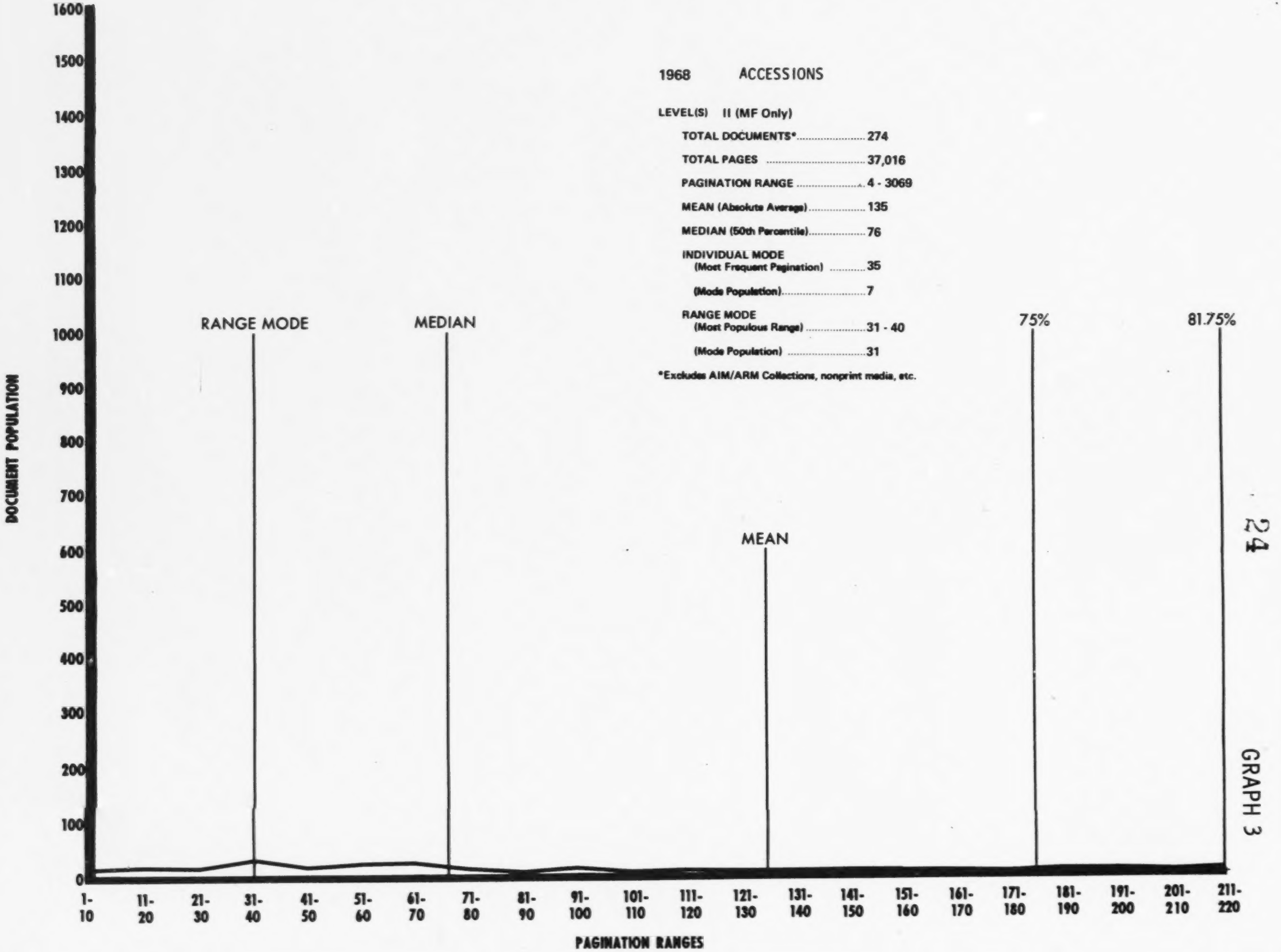
ERIC DATA BASE PAGINATION FIELD ANALYSIS

1968 ACCESSIONS

LEVEL(S) II (MF Only)

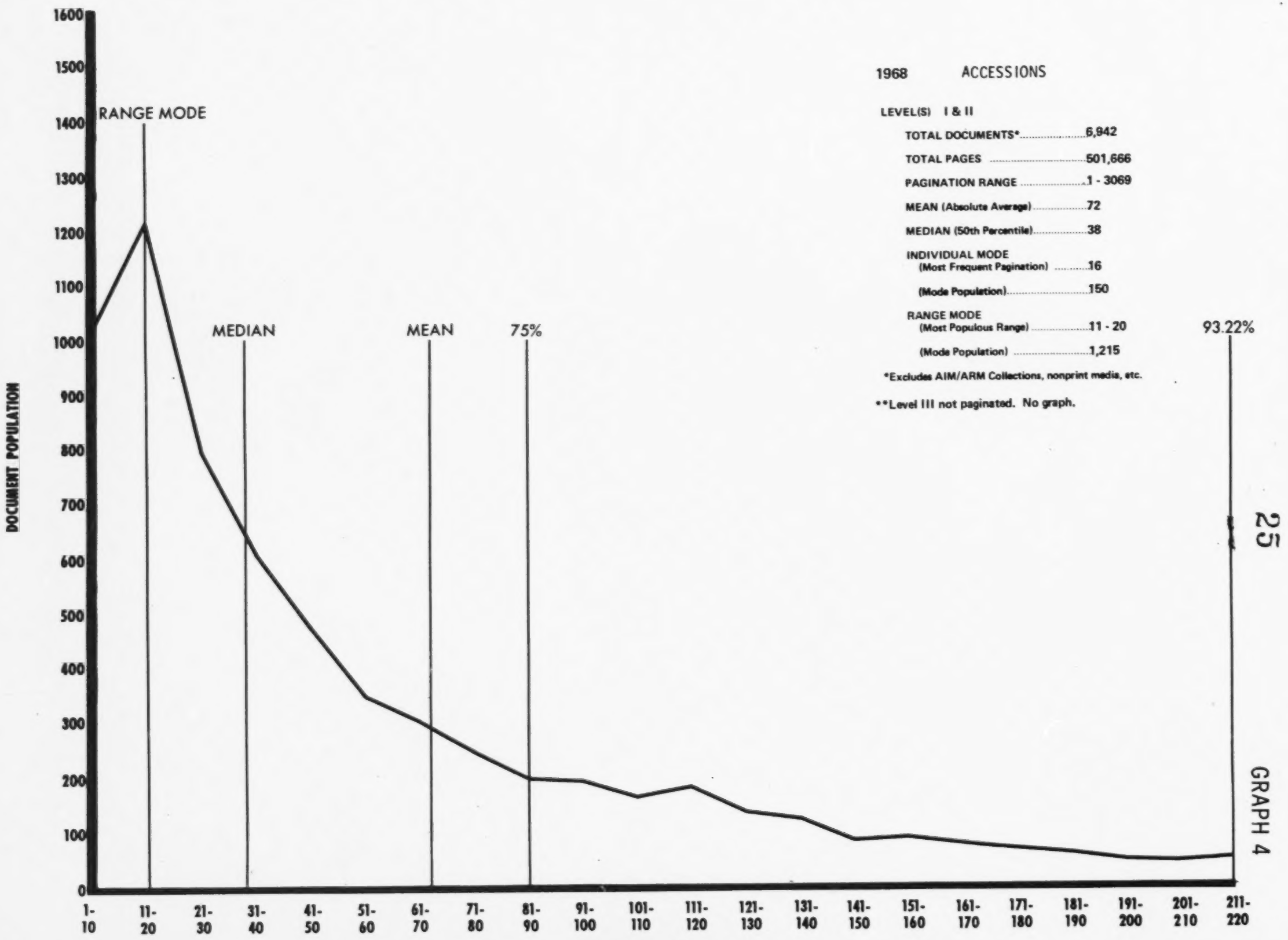
TOTAL DOCUMENTS*	274
TOTAL PAGES	37,016
PAGINATION RANGE	4 - 3069
MEAN (Absolute Average)	135
MEDIAN (50th Percentile)	76
INDIVIDUAL MODE (Most Frequent Pagination)	35
(Mode Population)	7
RANGE MODE (Most Populous Range)	31 - 40
(Mode Population)	31

*Excludes AIM/ARM Collections, nonprint media, etc.



GRAPH 3

ERIC DATA BASE PAGINATION FIELD ANALYSIS



1968 ACCESSIONS

LEVEL(S) I & II

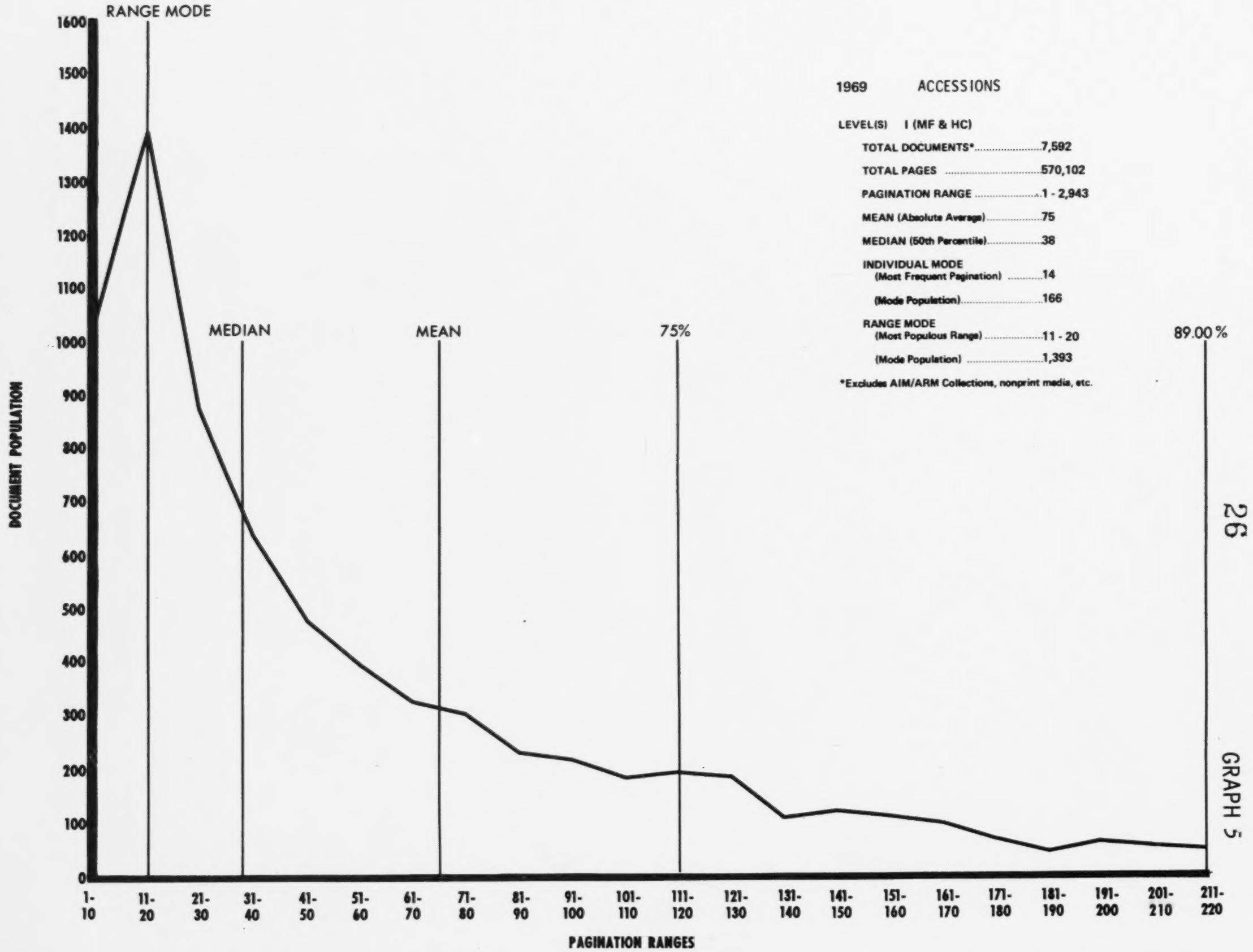
TOTAL DOCUMENTS*	6,942
TOTAL PAGES	501,666
PAGINATION RANGE	1 - 3069
MEAN (Absolute Average)	72
MEDIAN (50th Percentile)	38
INDIVIDUAL MODE (Most Frequent Pagination)	16
(Mode Population)	150
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	1,215

*Excludes AIM/ARM Collections, nonprint media, etc.

**Level III not paginated. No graph.

GRAPH 4

ERIC DATA BASE PAGINATION FIELD ANALYSIS



1969 ACCESSIONS

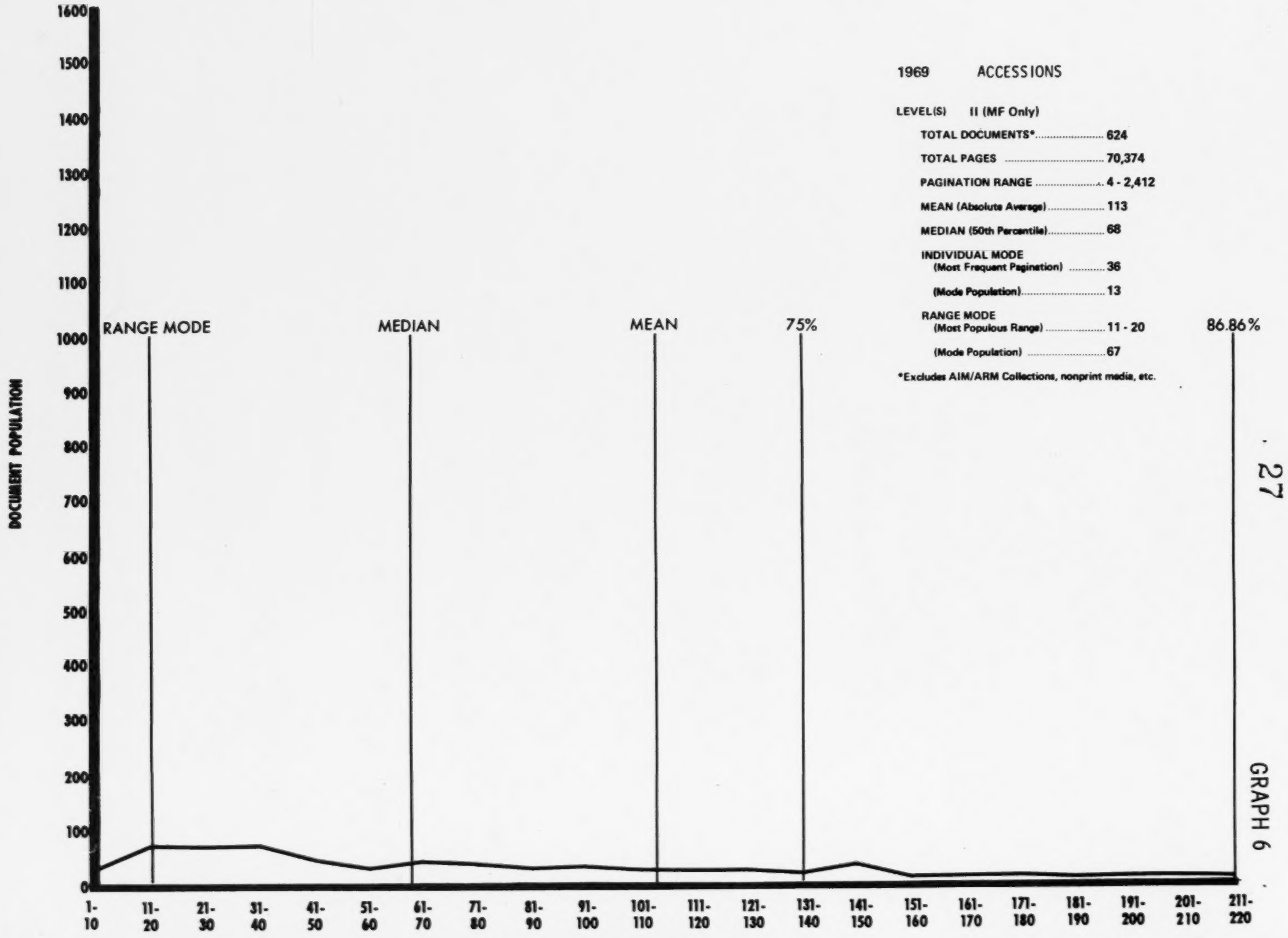
LEVEL(S) I (MF & HC)

TOTAL DOCUMENTS*	7,592
TOTAL PAGES	570,102
PAGINATION RANGE	1 - 2,943
MEAN (Absolute Average)	75
MEDIAN (50th Percentile)	38
INDIVIDUAL MODE (Most Frequent Pagination)	14
(Mode Population)	166
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	1,393

*Excludes AIM/ARM Collections, nonprint media, etc.

GRAPH 5

ERIC DATA BASE PAGINATION FIELD ANALYSIS



1969 ACCESSIONS

LEVEL(S) II (MF Only)

TOTAL DOCUMENTS*	624
TOTAL PAGES	70,374
PAGINATION RANGE	4 - 2,412
MEAN (Absolute Average)	113
MEDIAN (50th Percentile)	68
INDIVIDUAL MODE (Most Frequent Pagination)	36
(Mode Population)	13
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	67

*Excludes AIM/ARM Collections, nonprint media, etc.

GRAPH 6

ERIC DATA BASE PAGINATION FIELD ANALYSIS

1969 ACCESSIONS

LEVEL(S) III (Non-EDRS)

TOTAL DOCUMENTS*.....2,091

TOTAL PAGES 316,906

PAGINATION RANGE .....2 - 1,575

MEAN (Absolute Average).....152

MEDIAN (50th Percentile).....127

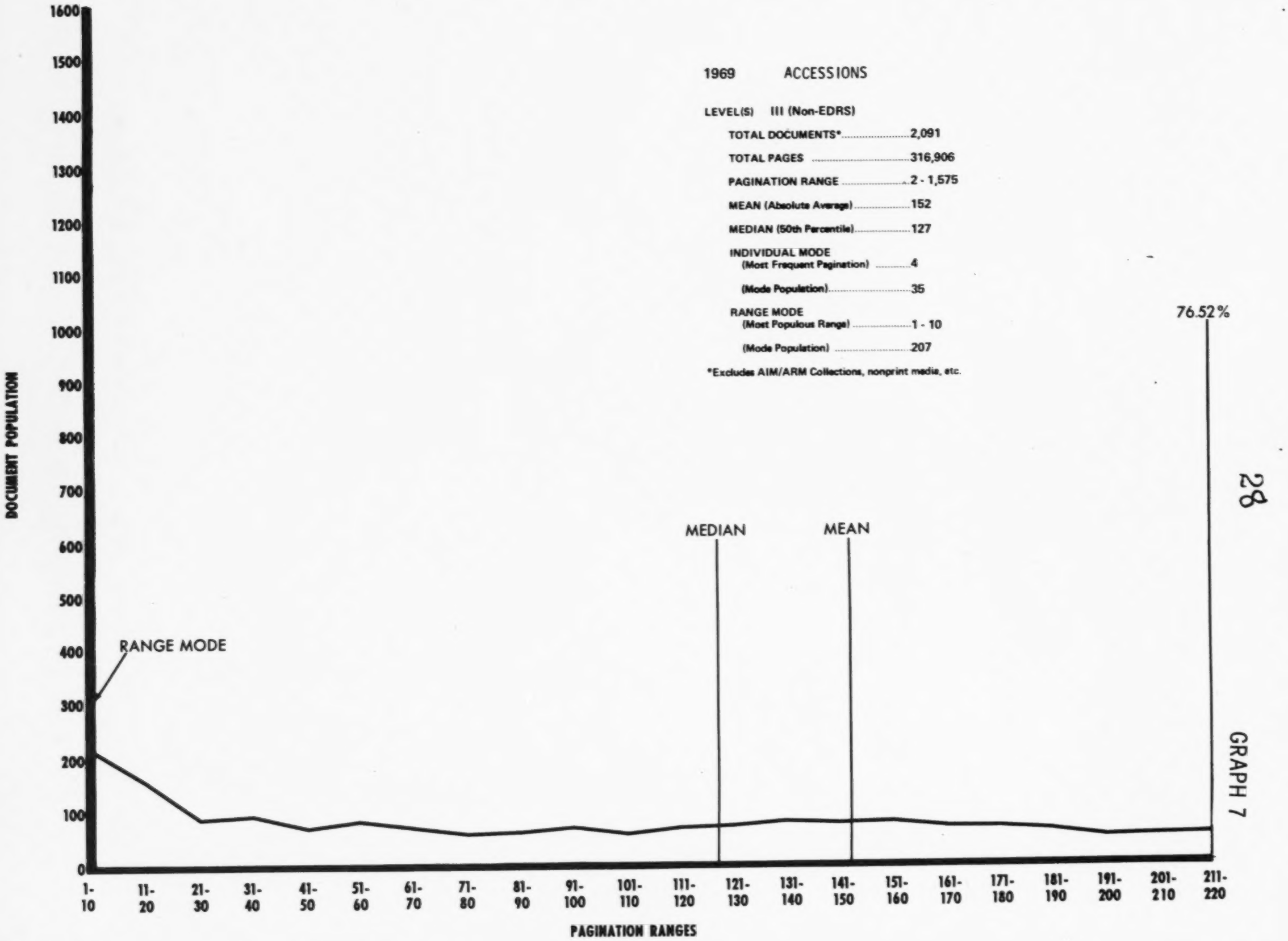
INDIVIDUAL MODE
(Most Frequent Pagination) 4

(Mode Population).....35

RANGE MODE
(Most Populous Range) .....1 - 10

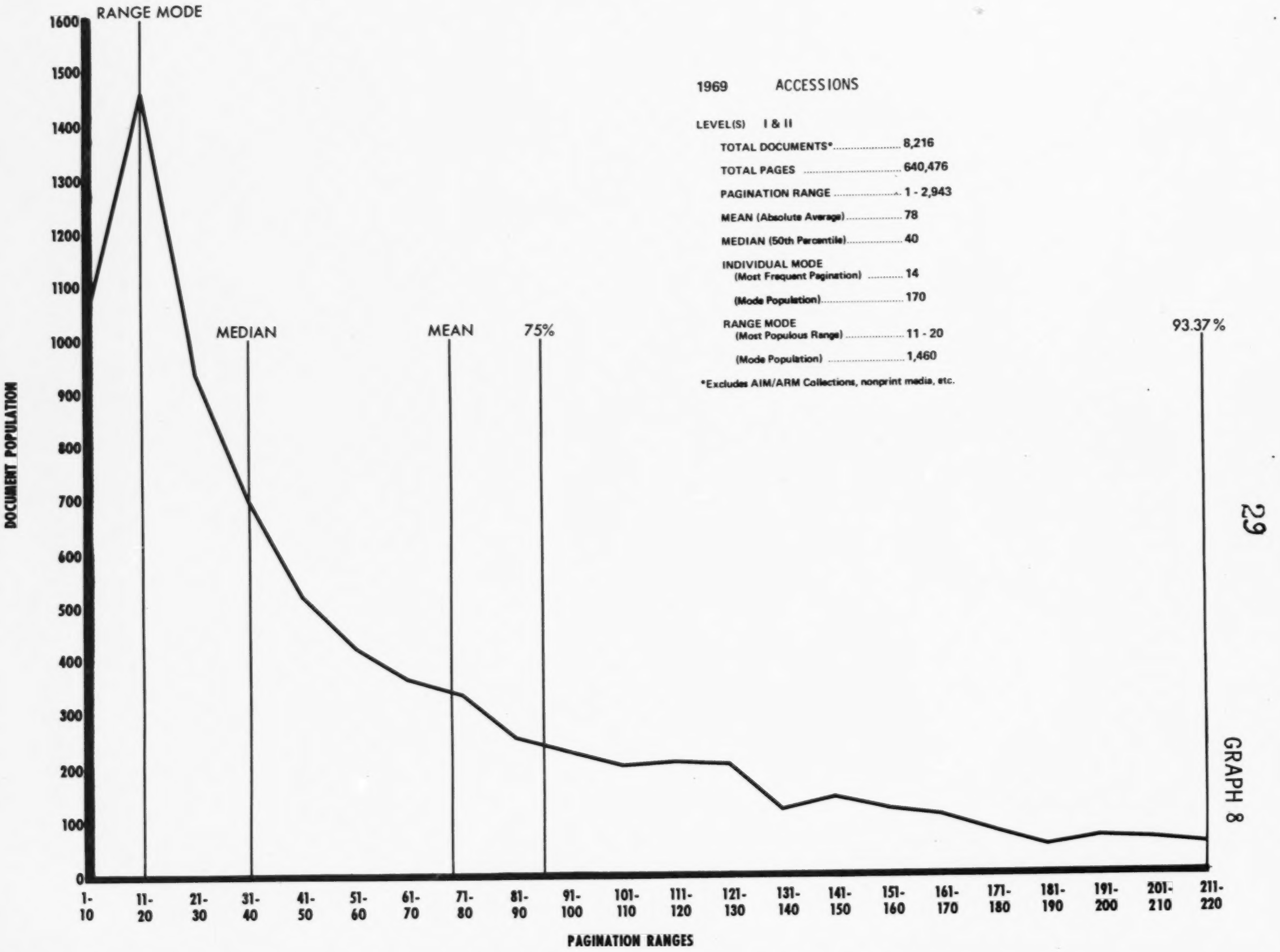
(Mode Population) 207

*Excludes AIM/ARM Collections, nonprint media, etc.



GRAPH 7

ERIC DATA BASE PAGINATION FIELD ANALYSIS



1969 ACCESSIONS

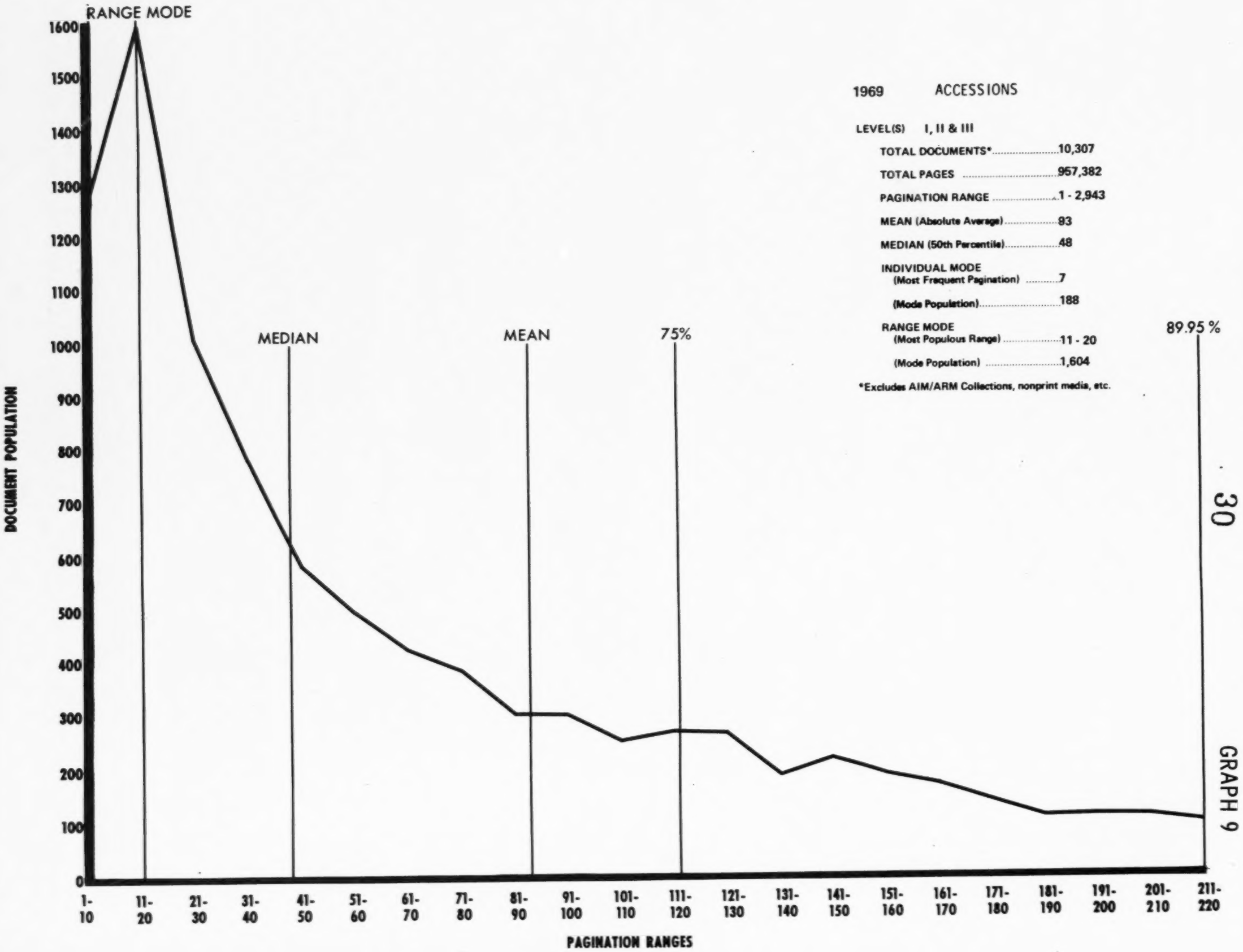
LEVEL(S) I & II

TOTAL DOCUMENTS*	8,216
TOTAL PAGES	640,476
PAGINATION RANGE	1 - 2,943
MEAN (Absolute Average)	78
MEDIAN (50th Percentile)	40
INDIVIDUAL MODE (Most Frequent Pagination)	14
(Mode Population)	170
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	1,460

*Excludes AIM/ARM Collections, nonprint media, etc.

GRAPH 8

ERIC DATA BASE PAGINATION FIELD ANALYSIS



1969 ACCESSIONS

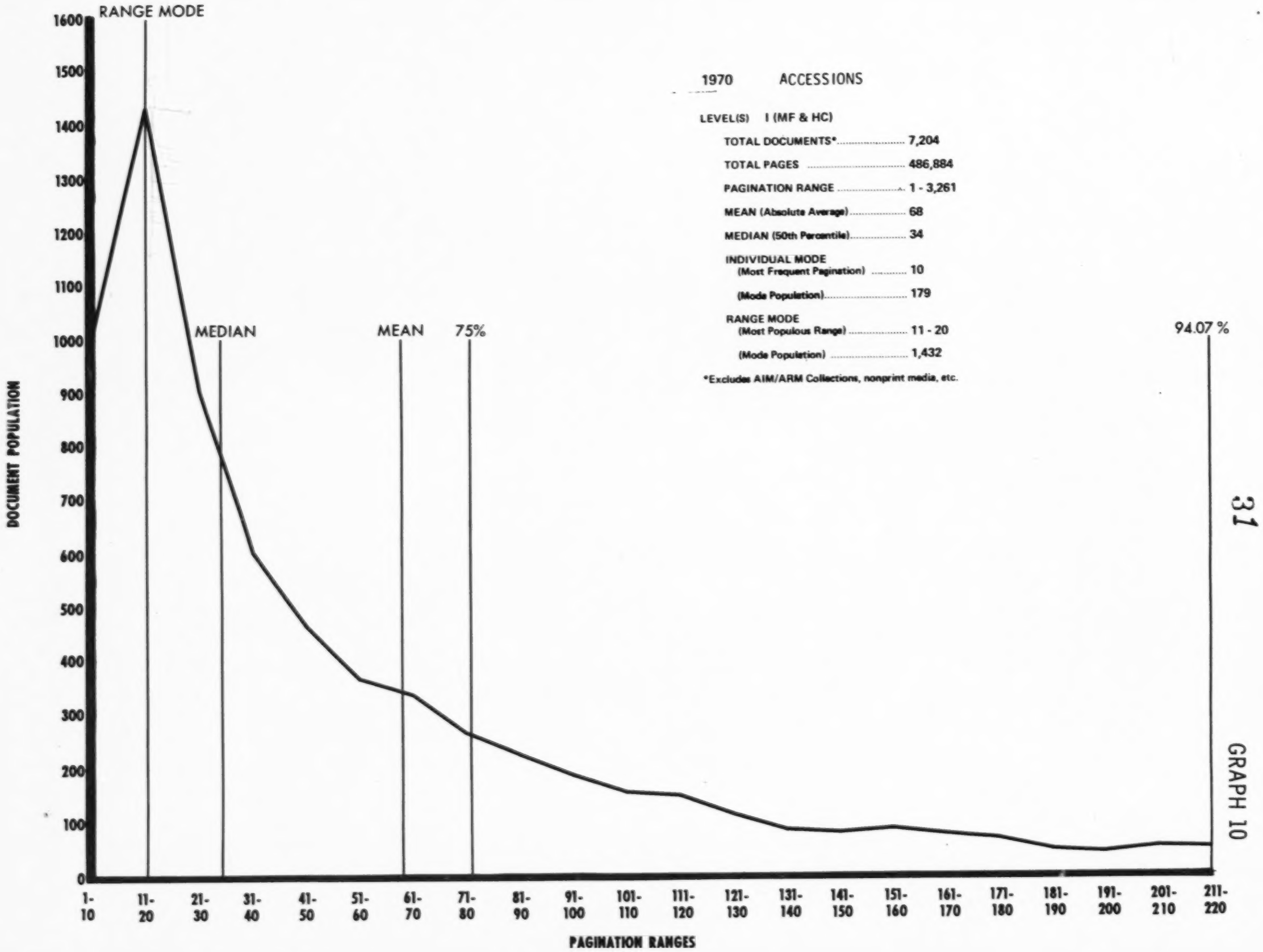
LEVEL(S) I, II & III

TOTAL DOCUMENTS*	10,307
TOTAL PAGES	957,382
PAGINATION RANGE	1 - 2,943
MEAN (Absolute Average)	93
MEDIAN (50th Percentile)	48
INDIVIDUAL MODE (Most Frequent Pagination)	7
(Mode Population)	188
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	1,604

*Excludes AIM/ARM Collections, nonprint media, etc.

GRAPH 9

ERIC DATA BASE PAGINATION FIELD ANALYSIS



1970 ACCESSIONS

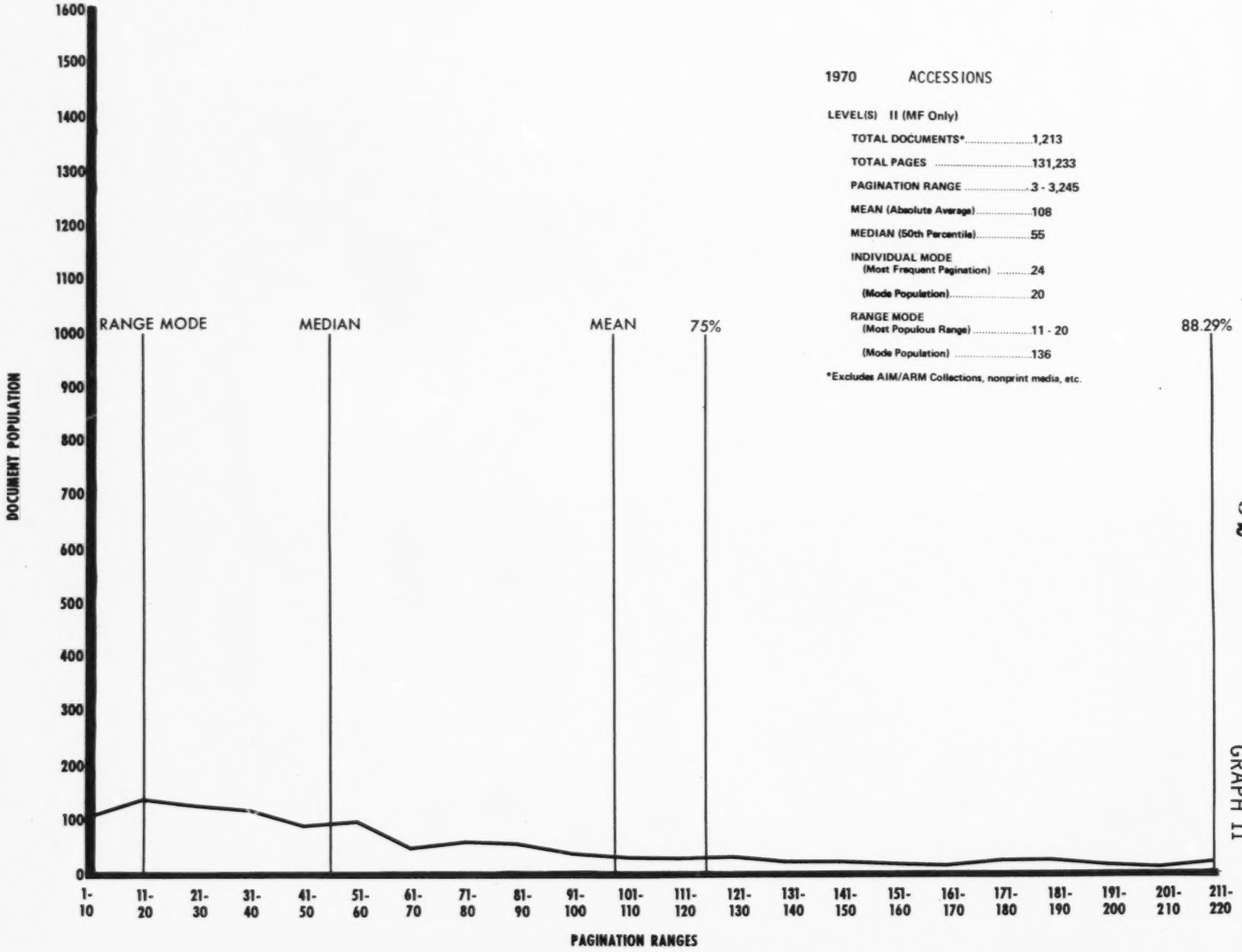
LEVEL(S) I (MF & HC)

TOTAL DOCUMENTS*	7,204
TOTAL PAGES	486,884
PAGINATION RANGE	1 - 3,261
MEAN (Absolute Average)	68
MEDIAN (50th Percentile)	34
INDIVIDUAL MODE (Most Frequent Pagination)	10
(Mode Population)	179
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	1,432

*Excludes AIM/ARM Collections, nonprint media, etc.

GRAPH 10

ERIC DATA BASE PAGINATION FIELD ANALYSIS



1970 ACCESSIONS

LEVEL(S) II (MF Only)

TOTAL DOCUMENTS*	1,213
TOTAL PAGES	131,233
PAGINATION RANGE	3 - 3,245
MEAN (Absolute Average)	108
MEDIAN (50th Percentile)	55
INDIVIDUAL MODE (Most Frequent Pagination)	24
(Mode Population)	20
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	136

*Excludes AIM/ARM Collections, nonprint media, etc.

GRAPH 11

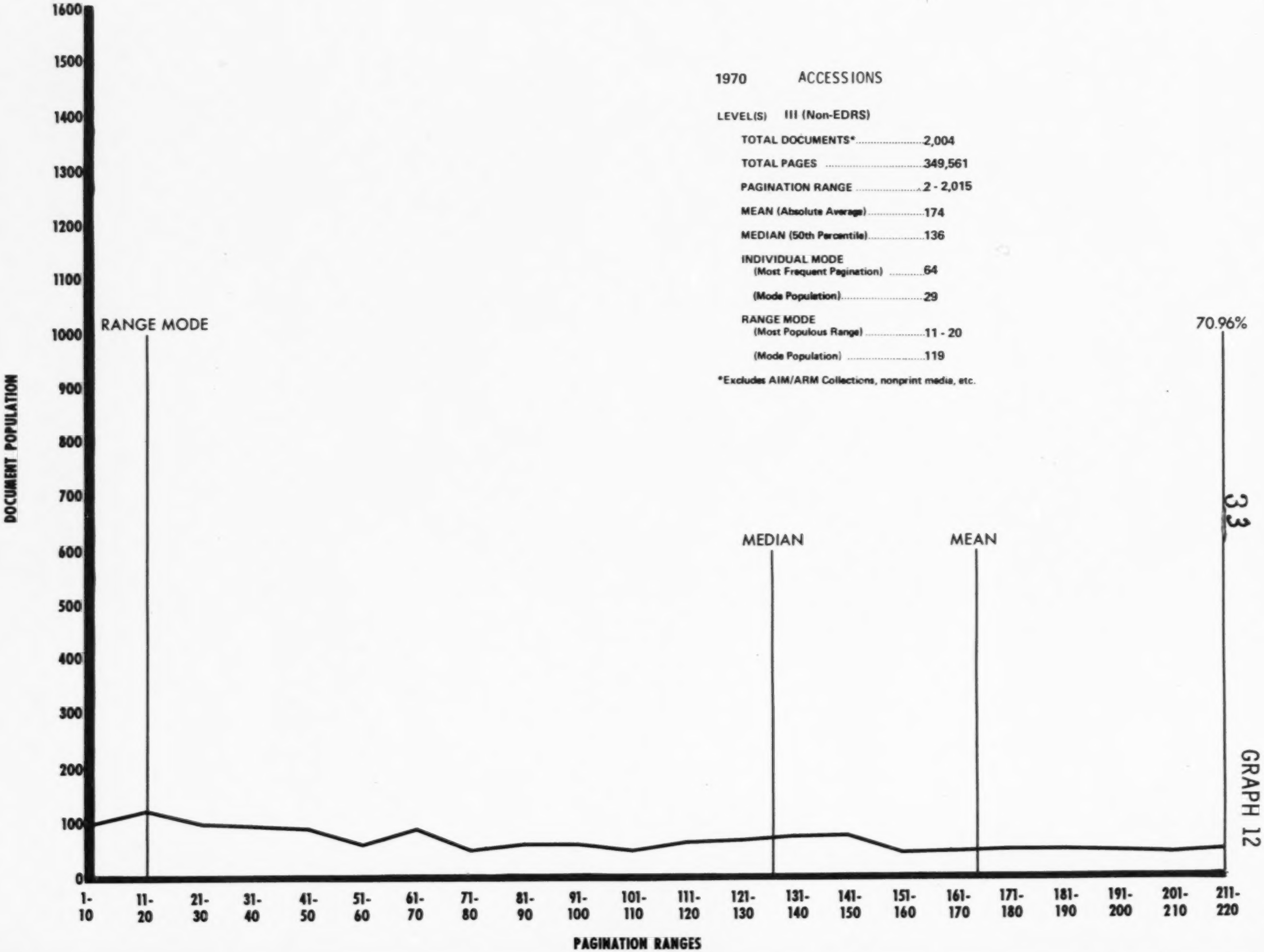
ERIC DATA BASE PAGINATION FIELD ANALYSIS

1970 ACCESSIONS

LEVEL(S) III (Non-EDRS)

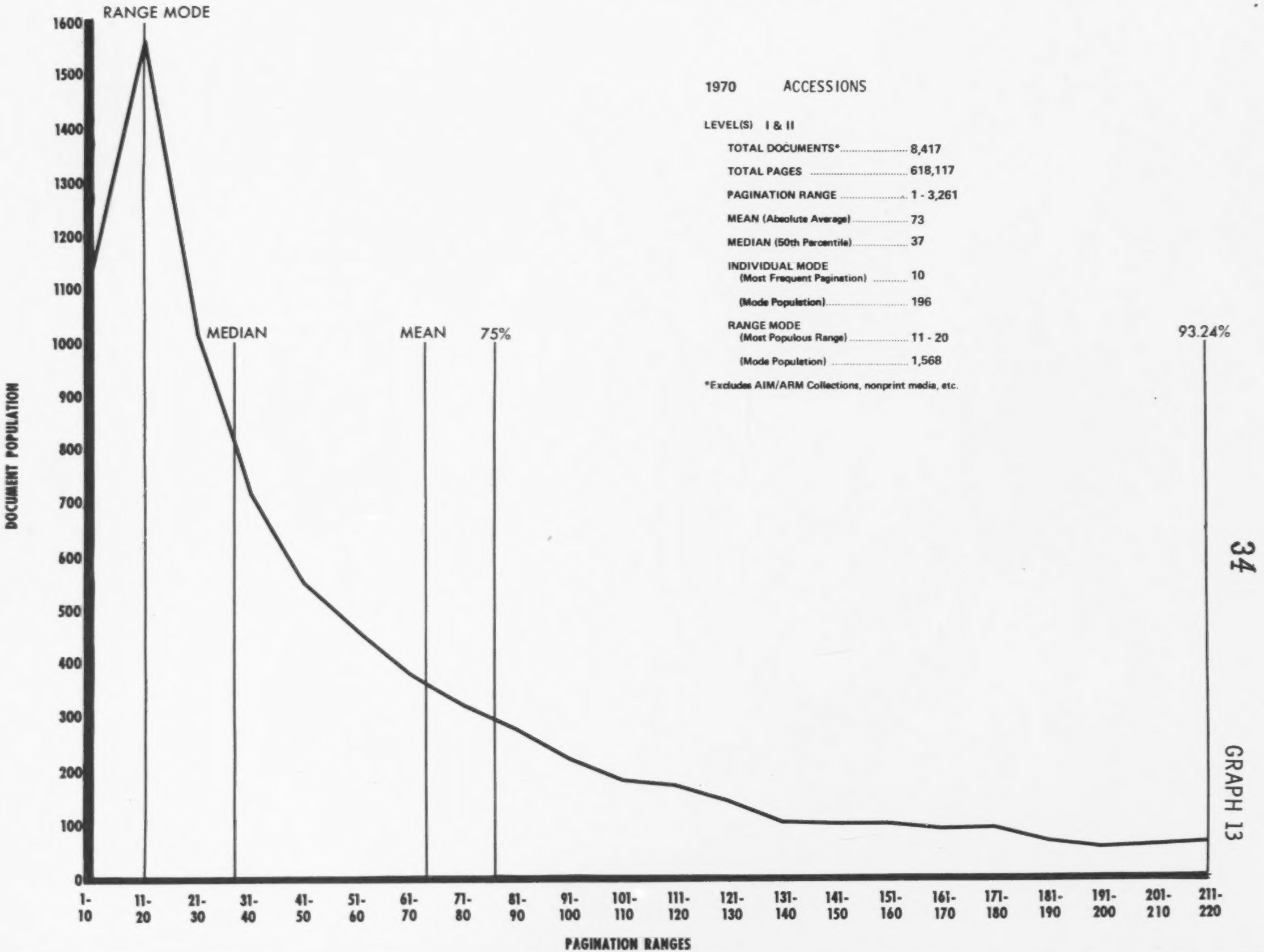
TOTAL DOCUMENTS*	2,004
TOTAL PAGES	349,561
PAGINATION RANGE	2 - 2,015
MEAN (Absolute Average)	174
MEDIAN (50th Percentile)	136
INDIVIDUAL MODE (Most Frequent Pagination)	64
(Mode Population)	29
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	119

*Excludes AIM/ARM Collections, nonprint media, etc.



GRAPH 12

ERIC DATA BASE PAGINATION FIELD ANALYSIS



1970 ACCESSIONS

LEVEL(S) I & II

TOTAL DOCUMENTS*	8,417
TOTAL PAGES	618,117
PAGINATION RANGE	1 - 3,261
MEAN (Absolute Average)	73
MEDIAN (50th Percentile)	37
INDIVIDUAL MODE (Most Frequent Pagination)	10
(Mode Population)	196
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	1,568

*Excludes AIM/ARM Collections, nonprint media, etc.

GRAPH 13

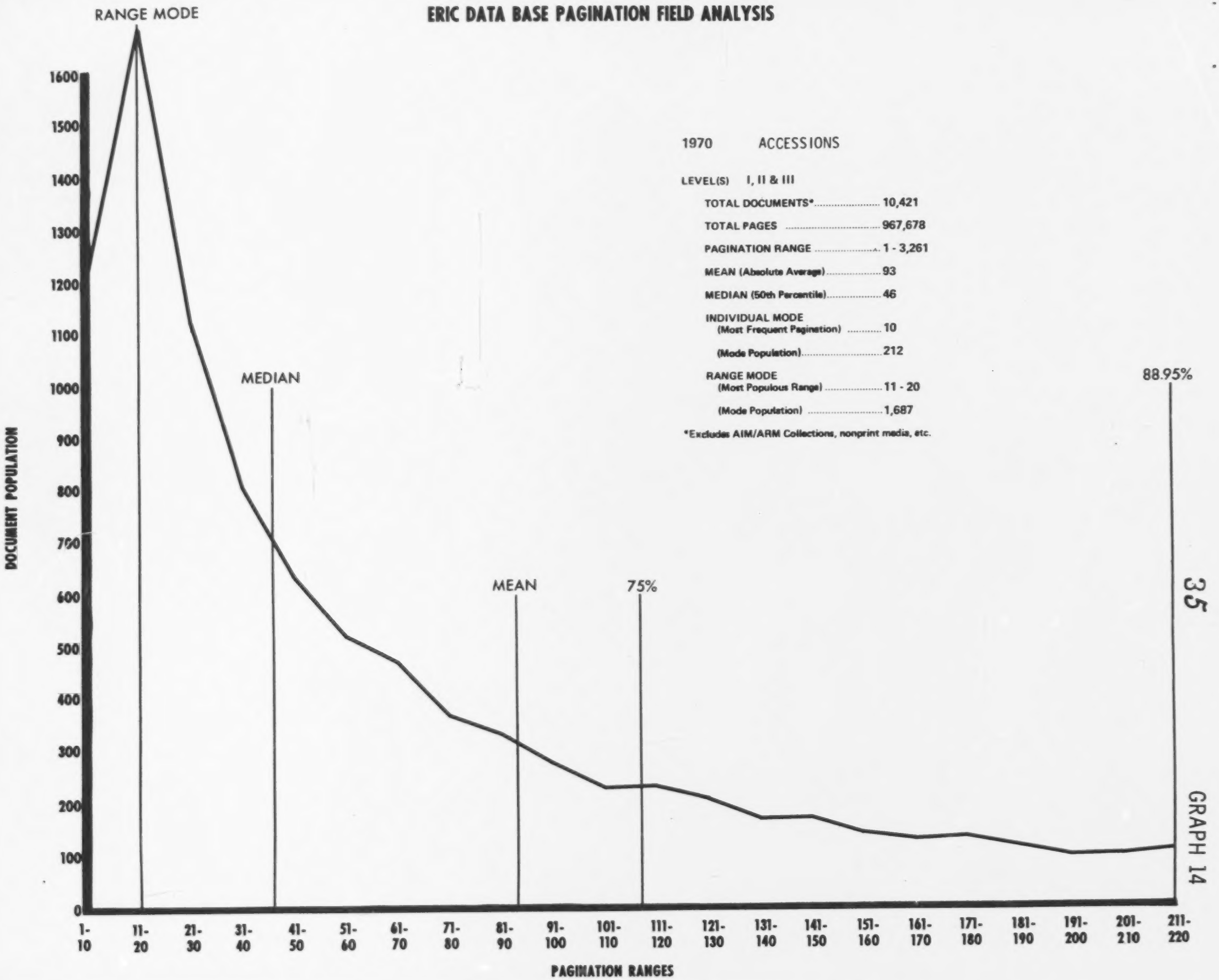
ERIC DATA BASE PAGINATION FIELD ANALYSIS

1970 ACCESSIONS

LEVEL(S) I, II & III

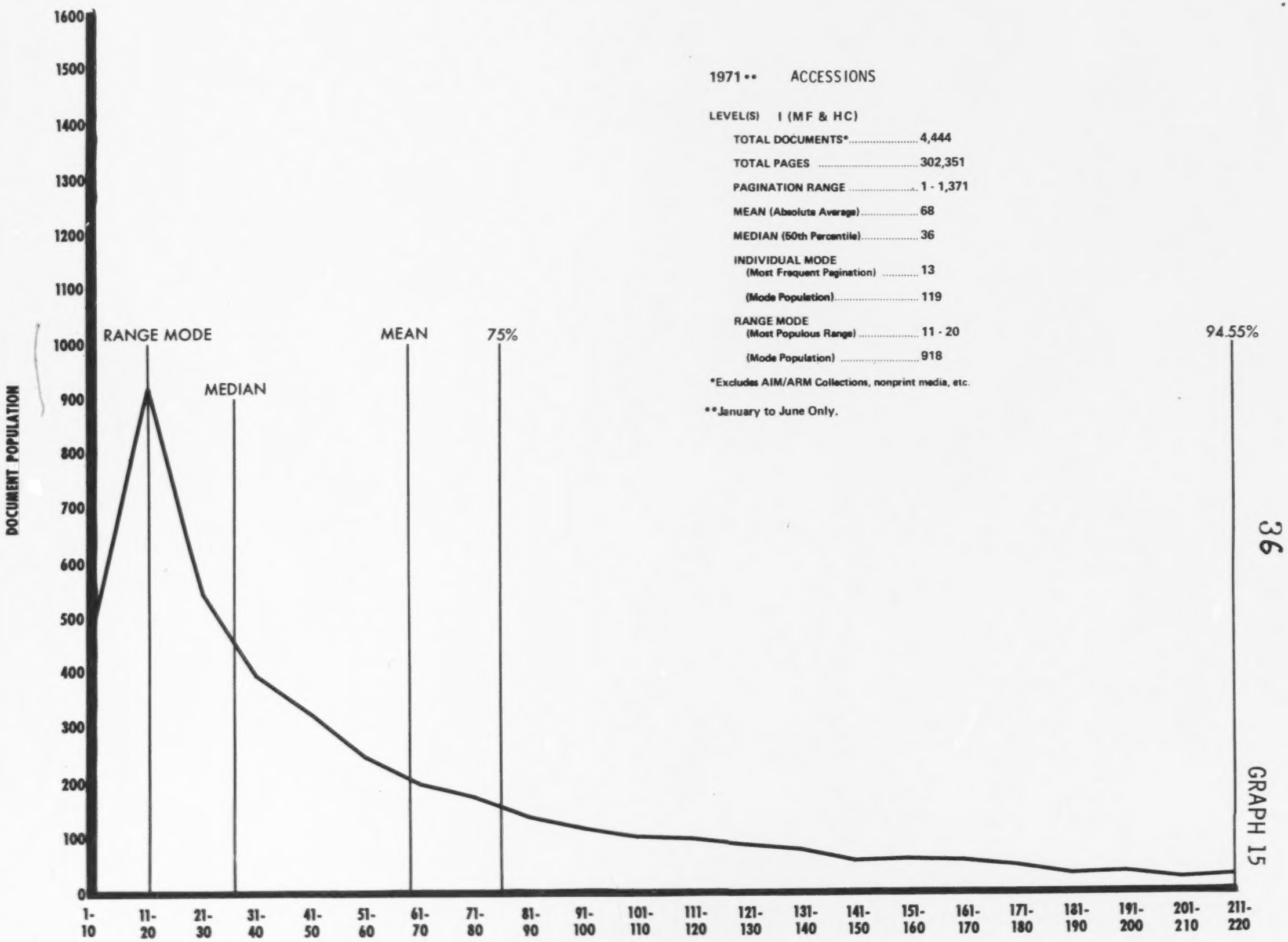
TOTAL DOCUMENTS*	10,421
TOTAL PAGES	967,678
PAGINATION RANGE	1 - 3,261
MEAN (Absolute Average)	93
MEDIAN (50th Percentile)	46
INDIVIDUAL MODE (Most Frequent Pagination)	10
(Mode Population)	212
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	1,687

*Excludes AIM/ARM Collections, nonprint media, etc.



GRAPH 14

ERIC DATA BASE PAGINATION FIELD ANALYSIS



1971 ** ACCESSIONS

LEVEL(S) I (MF & HC)

TOTAL DOCUMENTS*	4,444
TOTAL PAGES	302,351
PAGINATION RANGE	1 - 1,371
MEAN (Absolute Average)	68
MEDIAN (50th Percentile)	36
INDIVIDUAL MODE (Most Frequent Pagination)	13
(Mode Population)	119
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	918

*Excludes AIM/ARM Collections, nonprint media, etc.

**January to June Only.

36

GRAPH 15

ERIC DATA BASE PAGINATION FIELD ANALYSIS

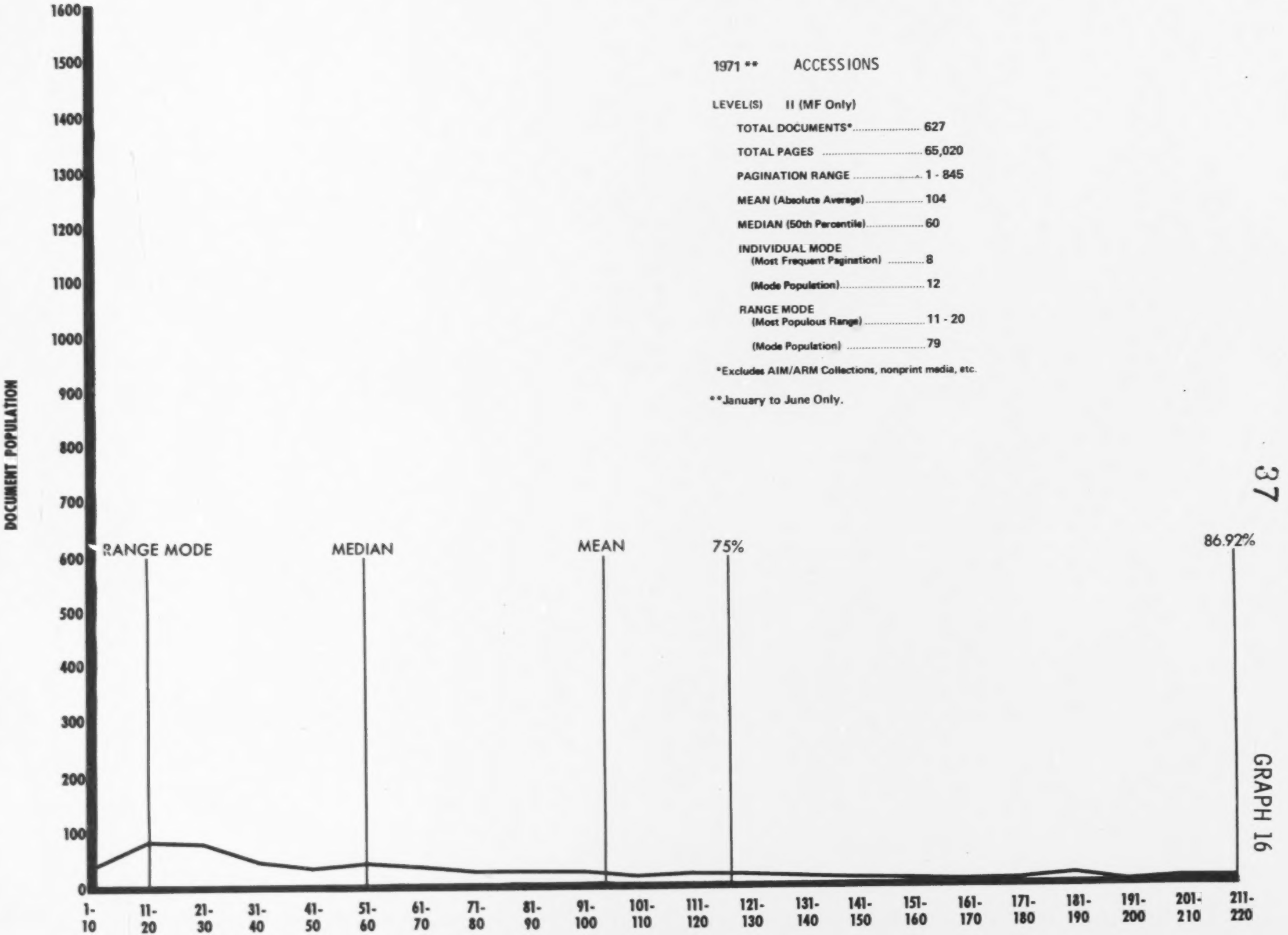
1971 ** ACCESSIONS

LEVEL(S) II (MF Only)

TOTAL DOCUMENTS*	627
TOTAL PAGES	65,020
PAGINATION RANGE	1 - 845
MEAN (Absolute Average)	104
MEDIAN (50th Percentile)	60
INDIVIDUAL MODE (Most Frequent Pagination)	8
(Mode Population)	12
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	79

*Excludes AIM/ARM Collections, nonprint media, etc.

**January to June Only.



37

GRAPH 16

ERIC DATA BASE PAGINATION FIELD ANALYSIS

1971 ** ACCESSIONS

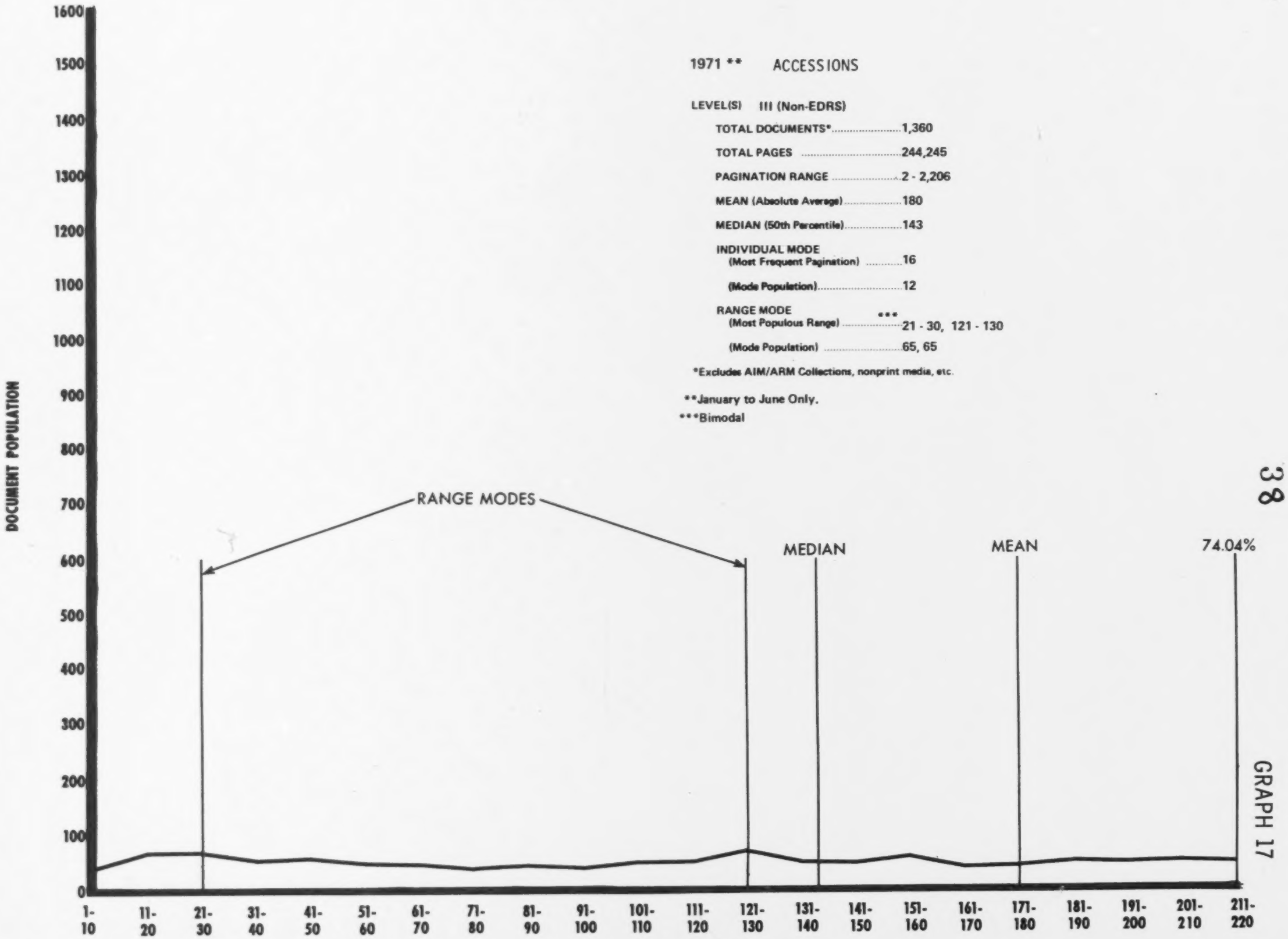
LEVEL(S) III (Non-EDRS)

TOTAL DOCUMENTS*	1,360
TOTAL PAGES	244,245
PAGINATION RANGE	2 - 2,206
MEAN (Absolute Average)	180
MEDIAN (50th Percentile)	143
INDIVIDUAL MODE (Most Frequent Pagination)	16
(Mode Population)	12
RANGE MODE (Most Populous Range) ***	21 - 30, 121 - 130
(Mode Population)	65, 65

*Excludes AIM/ARM Collections, nonprint media, etc.

**January to June Only.

***Bimodal



GRAPH 17

ERIC DATA BASE PAGINATION FIELD ANALYSIS

1971 ** ACCESSIONS

LEVEL(S) I & II

TOTAL DOCUMENTS*.....5,071

TOTAL PAGES 367,371

PAGINATION RANGE .....1 - 1,371

MEAN (Absolute Average).....72

MEDIAN (50th Percentile).....39

INDIVIDUAL MODE
(Most Frequent Pagination) 13

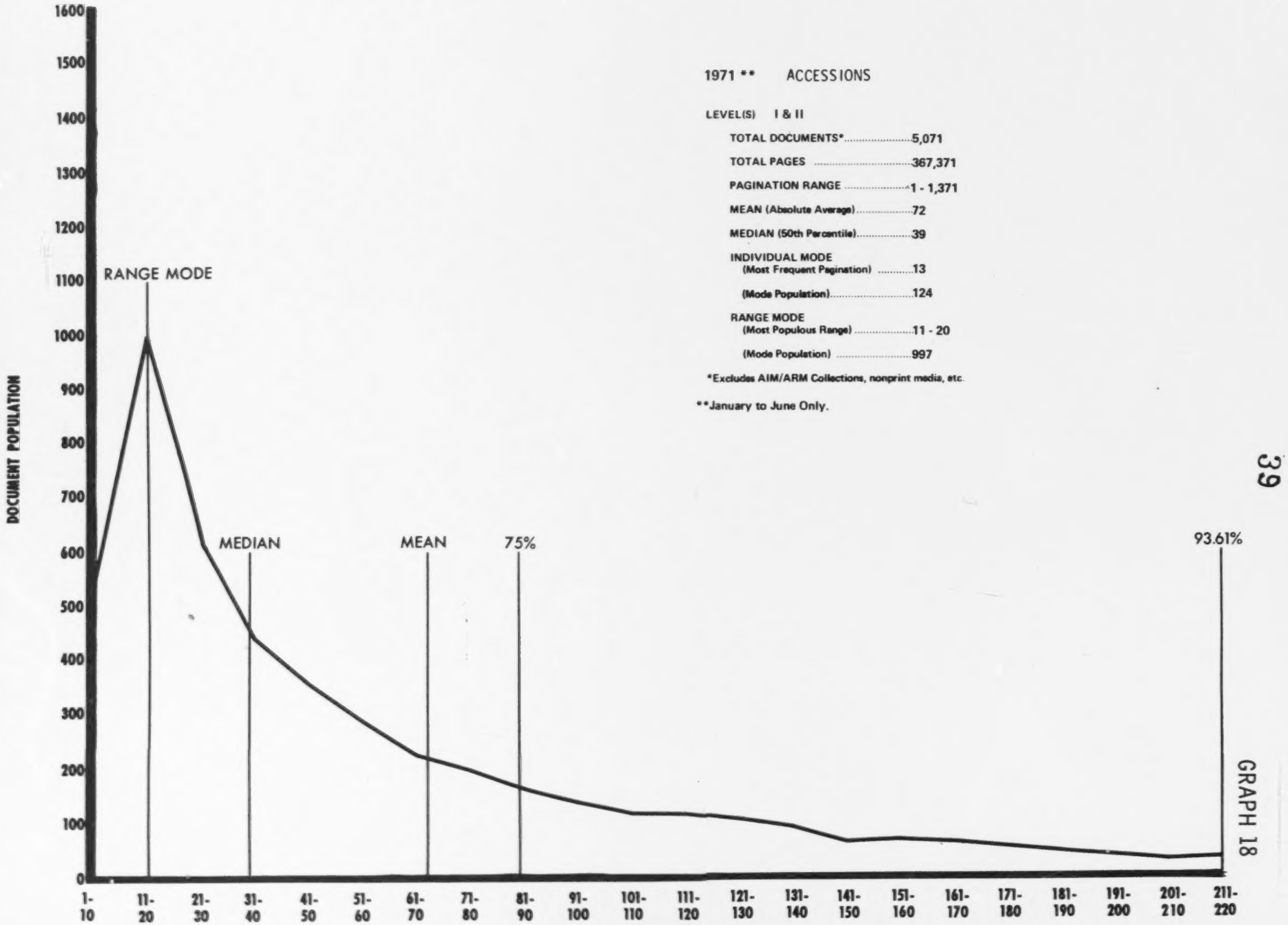
(Mode Population).....124

RANGE MODE
(Most Populous Range) .....11 - 20

(Mode Population) 997

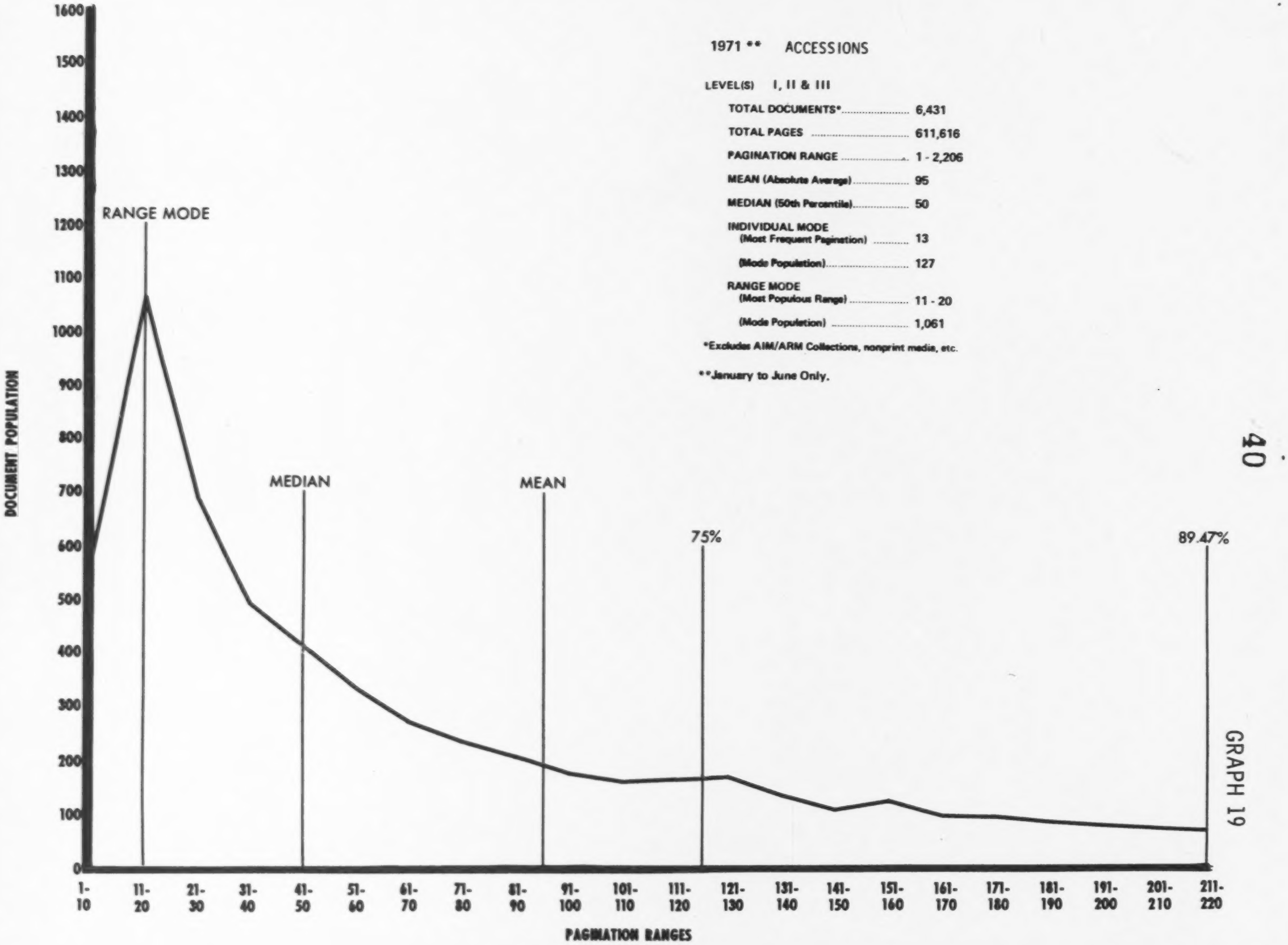
*Excludes AIM/ARM Collections, nonprint media, etc.

**January to June Only.



GRAPH 18

ERIC DATA BASE PAGINATION FIELD ANALYSIS



1971 ** ACCESSIONS

LEVEL(S) I, II & III

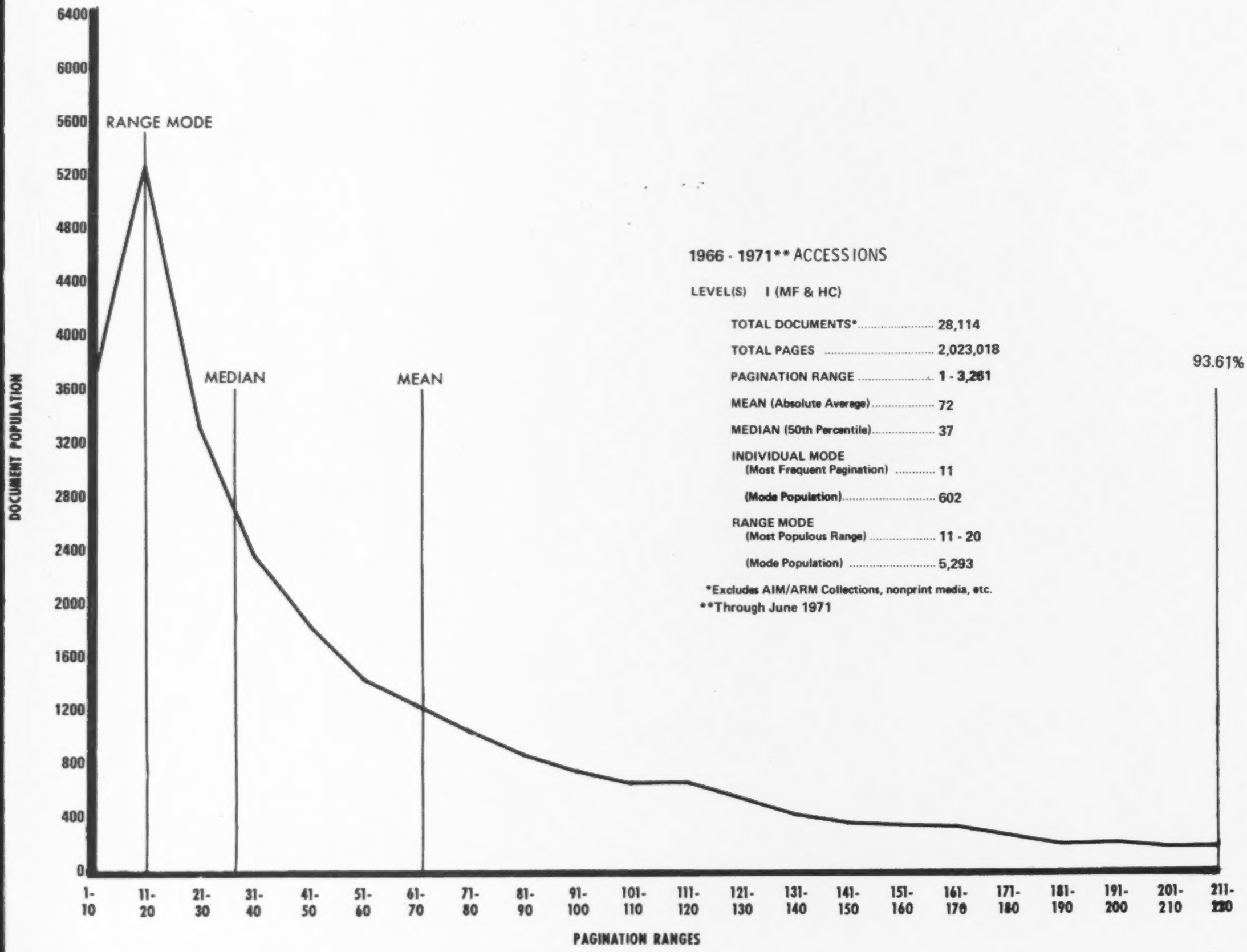
TOTAL DOCUMENTS*	6,431
TOTAL PAGES	611,616
PAGINATION RANGE	1 - 2,206
MEAN (Absolute Average)	95
MEDIAN (50th Percentile)	50
INDIVIDUAL MODE (Most Frequent Pagination)	13
(Mode Population)	127
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	1,061

*Excludes AIM/ARM Collections, nonprint media, etc.

**January to June Only.

GRAPH 19

ERIC DATA BASE PAGINATION FIELD ANALYSIS



1966 - 1971** ACCESSIONS

LEVEL(S) I (MF & HC)

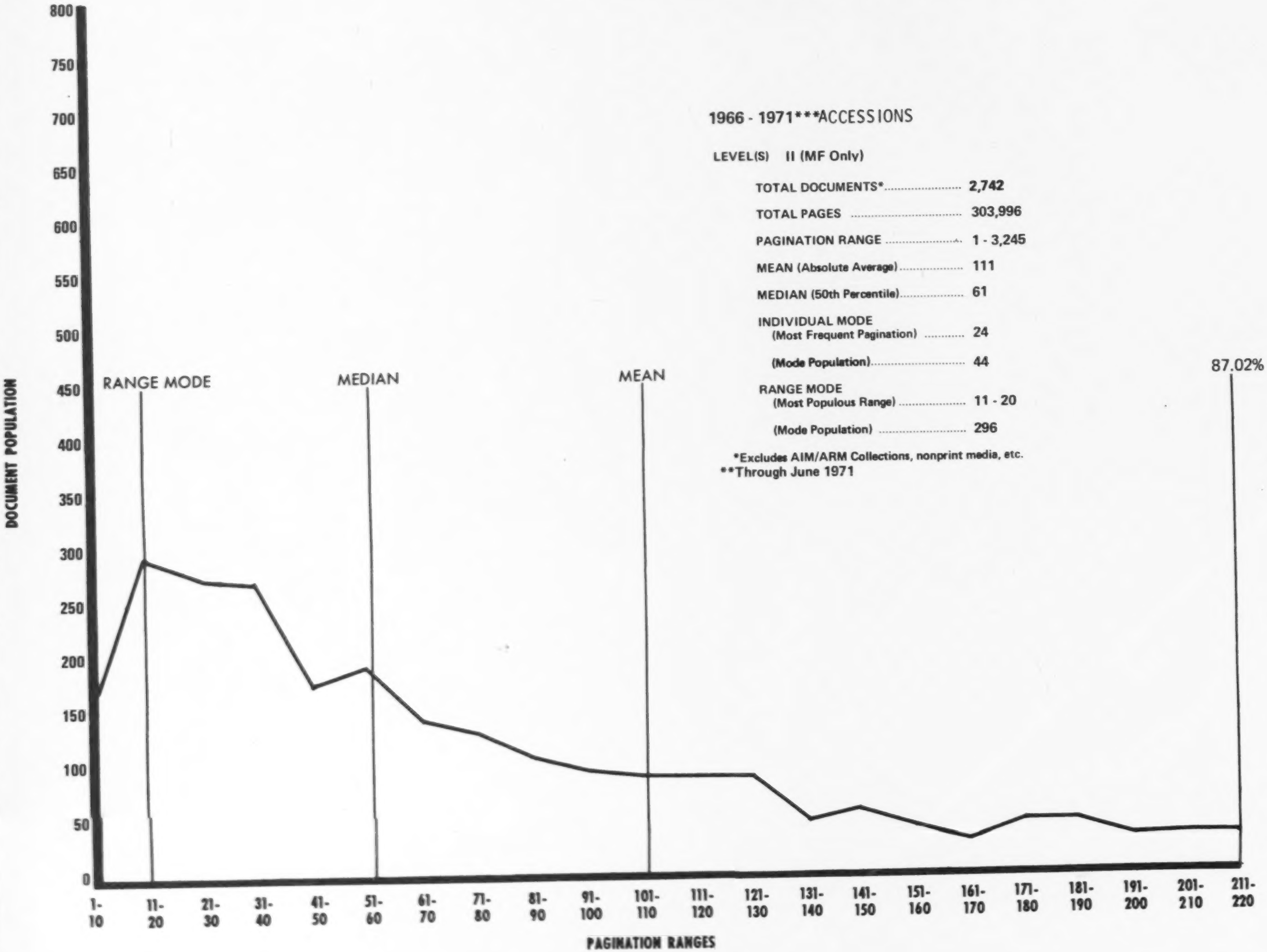
TOTAL DOCUMENTS*	28,114
TOTAL PAGES	2,023,018
PAGINATION RANGE	1 - 3,261
MEAN (Absolute Average)	72
MEDIAN (50th Percentile)	37
INDIVIDUAL MODE (Most Frequent Pagination)	11
(Mode Population)	602
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	5,293

*Excludes AIM/ARM Collections, nonprint media, etc.

**Through June 1971

GRAPH 20

ERIC DATA BASE PAGINATION FIELD ANALYSIS



1966 - 1971***ACCESSIONS

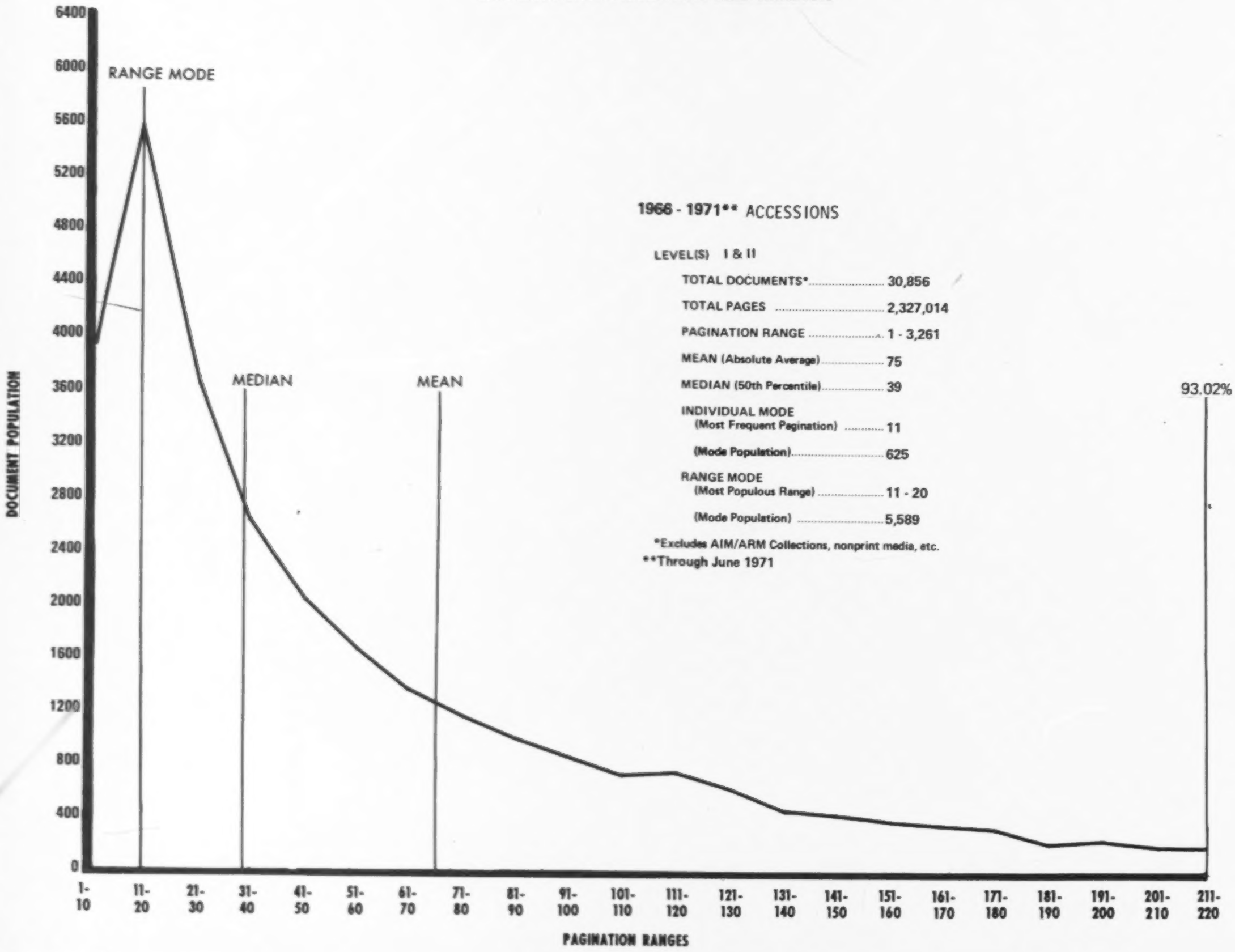
LEVEL(S) II (MF Only)

TOTAL DOCUMENTS*	2,742
TOTAL PAGES	303,996
PAGINATION RANGE	1 - 3,245
MEAN (Absolute Average)	111
MEDIAN (50th Percentile)	61
INDIVIDUAL MODE (Most Frequent Pagination)	24
(Mode Population)	44
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	296

*Excludes AIM/ARM Collections, nonprint media, etc.
**Through June 1971

GRAPH 21

ERIC DATA BASE PAGINATION FIELD ANALYSIS



1966 - 1971** ACCESSIONS

LEVEL(S) I & II

TOTAL DOCUMENTS*	30,856
TOTAL PAGES	2,327,014
PAGINATION RANGE	1 - 3,261
MEAN (Absolute Average)	75
MEDIAN (50th Percentile)	39
INDIVIDUAL MODE (Most Frequent Pagination)	11
(Mode Population)	625
RANGE MODE (Most Populous Range)	11 - 20
(Mode Population)	5,589

*Excludes AIM/ARM Collections, nonprint media, etc.

**Through June 1971

GRAPH 22

ERIC DATA BASE PAGINATION FIELD ANALYSIS

1969-70 ACCESSIONS

LEVEL(S) I, II & III

TOTAL DOCUMENTS*	20,732
TOTAL PAGES	1,925,060
PAGINATION RANGE	1-3,261
MEAN (Absolute Average)	93
MEDIAN (50th Percentile)	47
INDIVIDUAL MODE (Most Frequent Pagination)	14
(Mode Population)	386
RANGE MODE (Most Populous Range)	11-20
(Mode Population)	3,291

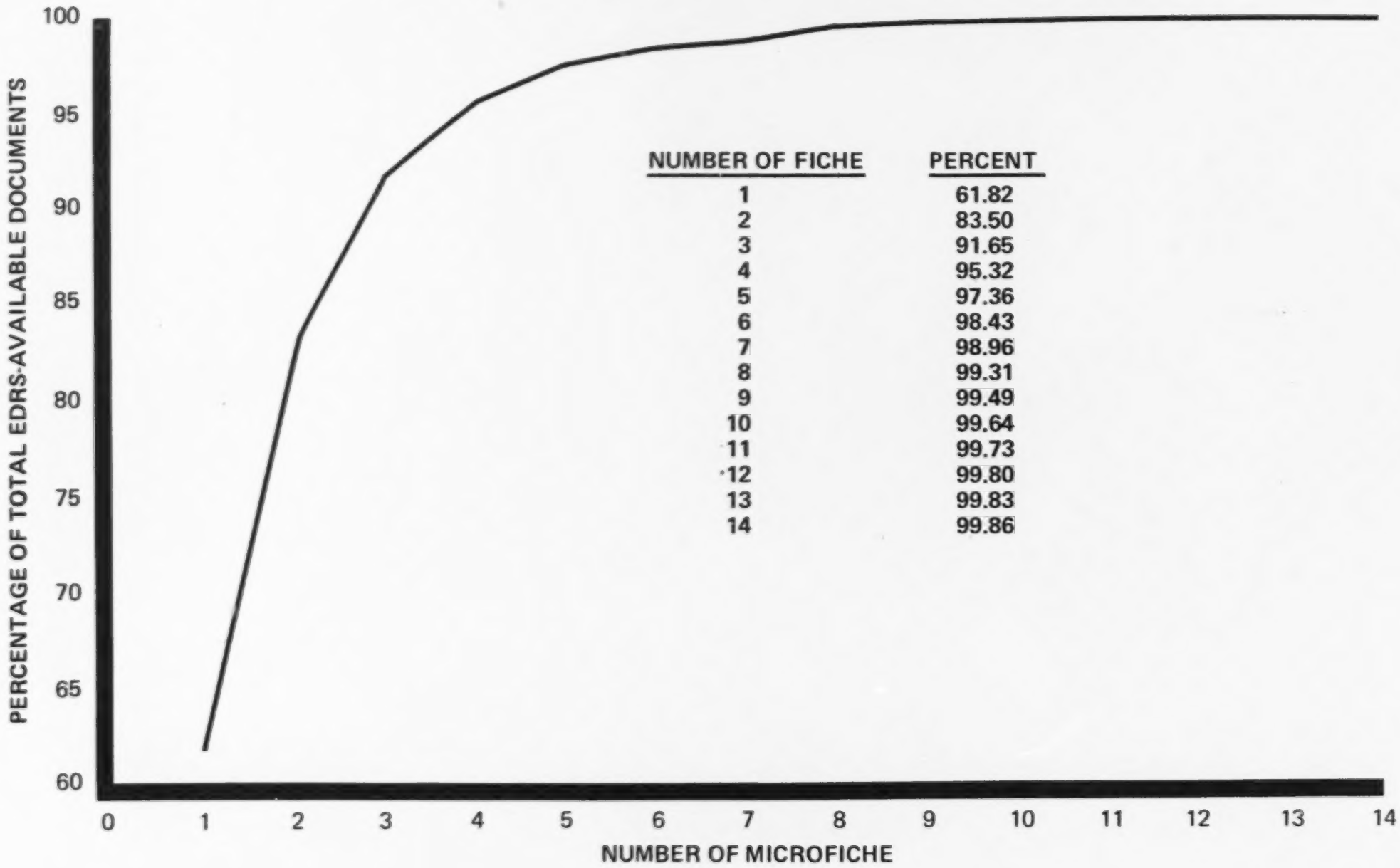
*Excludes AIM/ARM Collections, nonprint media, etc.

GRAPH 23

ERIC DATA BASE

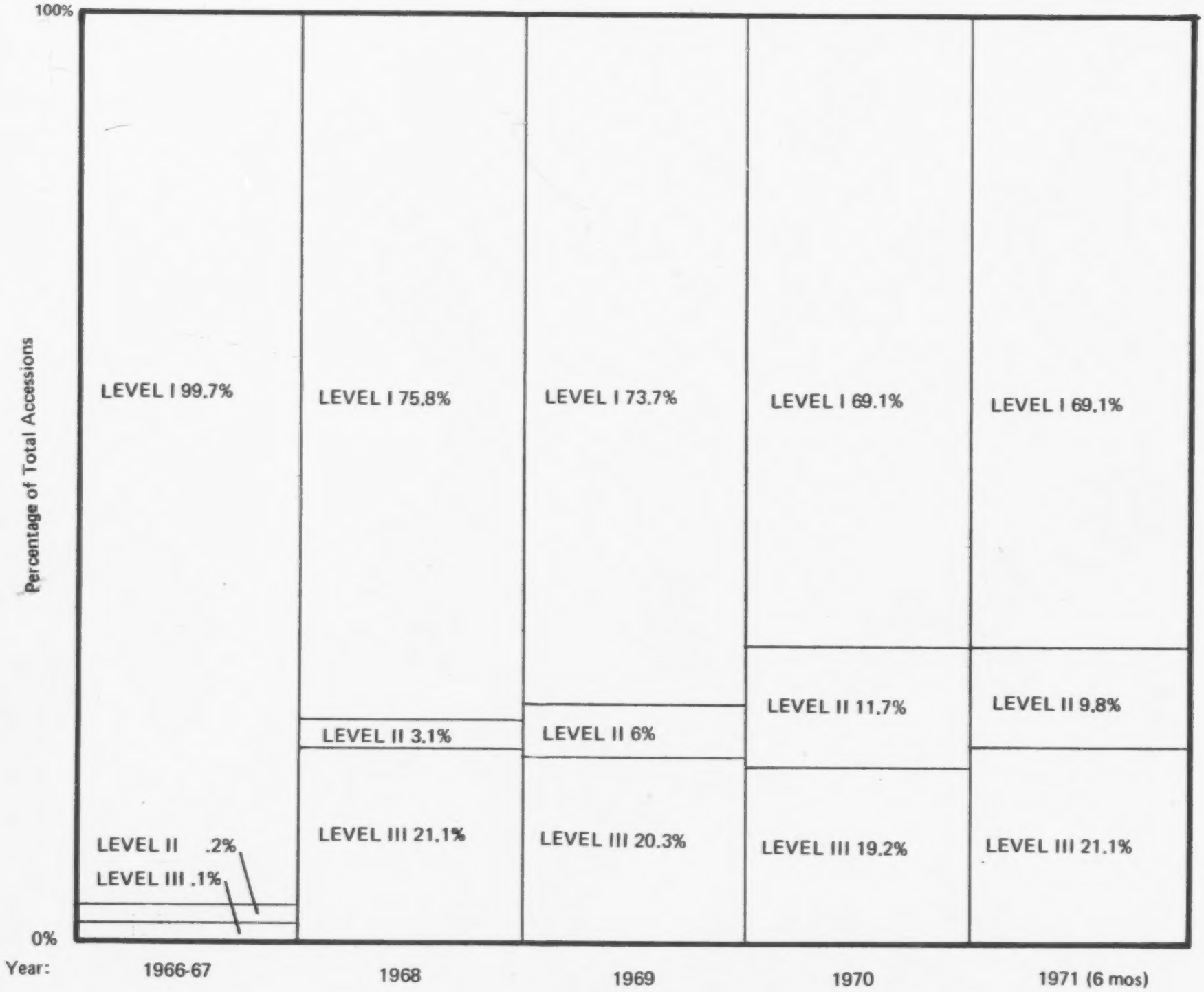
PAGINATION FIELD ANALYSIS

MICROFICHE REQUIRED FOR EDRS-AVAILABLE DOCUMENTS (LEVELS I AND II)



NUMBER OF FICHE	PERCENT
1	61.82
2	83.50
3	91.65
4	95.32
5	97.36
6	98.43
7	98.96
8	99.31
9	99.49
10	99.64
11	99.73
12	99.80
13	99.83
14	99.86

GRAPH 24



ERIC Data Base (ED Accessions)
Total Accessions Subdivided by Availability Level (I, II, III)

GRAPH 25